

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 16:56:51 ; Search time 70.84 Seconds
(without alignments)
315.918 Million cell updates/sec

Title: US-09-490-700-36
Perfect score: 840
Sequence: 1 MVTPTWMDNPIEVVNDV.....ECVANSVILQNNFEFTIID 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	840	100.0	153	4	O95280 homo sapien
2	840	100.0	153	12	O42043 unclassified
3	831	98.9	560	12	O9C06 human endog
4	811	96.5	153	4	O9UNW3 homo sapien
5	805	95.8	153	4	O9UNW2 homo sapien
6	805	95.8	245	4	O9HDB8 homo sapien
7	792	94.3	1361	4	O14273 homo sapien
8	789	93.9	2294	4	O9UKH9 homo sapien
9	784	93.3	153	4	O95284 homo sapien
10	782	93.1	153	4	O95281 homo sapien
11	779	92.7	694	12	O9INA9 human endog
12	779	92.7	699	4	O9UBU4 homo sapien
13	779	92.7	699	12	O93384 human endog
14	778	92.6	514	12	O69386 human endog
15	778	92.6	694	12	O9YNA7 human endog
16	777	92.5	153	4	O95282 homo sapien
17	777	92.5	694	12	O9YNA5 human endog
18	774	92.1	277	4	O9UKH7 homo sapien
19	752.5	89.6	698	4	O9UKH3 homo sapien

ALIGNMENTS

RESULT 1

O95280 ID O95280 PRELIMINARY; PRT; 153 AA.
AC O95280;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE ENVELOPE PROTEIN RIC-1.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449692; PubMed=9778243;
RA Lan M.S., Mason A., Coutant R., Chen Q.Y., Vargas A., Rao J.,
RA Gomez R., Chalew S., Garry R., MacLaren N.K.;
RT "HERV-K10s and immune-mediated (type 1) diabetes.";
RL Cell 95:14-16(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Hasuike S., Jinno Y.;
RT "Isolation and localization of an endogenous retrovirus gene, a
RT candidate gene for type 1 diabetes, and identification of a CA repeat
RT marker at its locus."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF084864; AAC68893.1; -;
DR EMBL; AF134984; AAD33055.1; -;
KW Envelope protein.
SQ SEQUENCE 153 AA; 17305 MW; 2B36E1656DF0C01A CRC64;

Query Match 100.0%; Score 840; DB 4; Length 153;

Best Local Similarity 100.0%; Pred. No. 1.4e-81;

Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVTPTWMDNPIEVVNDVSYVWPGTDDRCAPKPEEGMMINISIGVHYPPICLGRAPGC 60

Db 1 MVTPTWMDNPIEVVNDVSYVWPGTDDRCAPKPEEGMMINISIGVHYPPICLGRAPGC 60

QY 61 LMPAVQNLWEVPTVSPNSRFTYHMSGLSRPRVNYLQDFSYQSRSLKFRPKGKTCPKEI 120

|||||
Db 61 LMPAVQNNLVVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPKREI 120
Qy 121 PGSKNTEVLWEECVANSVVLQNNFEFTIID 153
Db 121 PGSKNTEVLWEECVANSVVLQNNFEFTIID 153
RESULT 2
O42043
ID O42043 PRELIMINARY; PRT; 153 AA.
AC O42043;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE ENVELOPE PROTEIN.
GN ENV.
OS unclassified Retroviridae.
OC Viruses; Retroviral viruses; Retroviridae.
OX NCBI_TaxID=35276;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97386455; PubMed=9244304;
RA Conrad R., Weissmahr R.N., Boni J., Arcari R., Schubach J., Mach B.
RT "A human endogenous retroviral superantigen as candidate autoantigen
gene in type 1 diabetes.";
RL Cell 90:303-313(1997).
DR EMBL; AF012337; AAC58457.1; -;
DR EMBL; AF012336; AAC58456.1; -;
KW Envelope protein.
SQ SEQUENCE 153 AA; 17305 MW; 2B36E1656DF0C01A CRC64;

inventions

7-25-97

Query Match 100.0%; Score 840; DB 12; Length 153;
Best Local Similarity 100.0%; Pred. No. 1.4e-81;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MVTPTWMDNPIEVYVNDVSVVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60
Db 1 MVTPTWMDNPIEVYVNDVSVVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60
Qy 61 LMPAVQNNLVVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPKREI 120
Db 61 LMPAVQNNLVVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPKREI 120
Qy 121 PGSKNTEVLWEECVANSVVLQNNFEFTIID 153
Db 121 PGSKNTEVLWEECVANSVVLQNNFEFTIID 153

RESULT 3
O9QC06
ID O9QC06 PRELIMINARY; PRT; 560 AA.
AC O9QC06;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE ENV PROTEIN.
GN ENV.
OS Human endogenous retrovirus K.
OC Viruses; Retroviral viruses; Retroviridae.
OX NCBI_TaxID=45617;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99445825; PubMed=10516026;
RA Toenjes R.R., Czauderna F., Kurth R.;
RT "Genome wide screening, cloning, chromosomal assignment and expression
of full-length human endogenous retrovirus type K (HERV-K).";
RL J. Virol. 73:9187-9195(1999).
DR EMBL; Y18890; CAB56604.1; -;
SQ SEQUENCE 560 AA; 63671 MW; 8A4565663901BC3A CRC64;

Query Match 98.9%; Score 831; DB 12; Length 560;
Best Local Similarity 99.3%; Pred. No. 5.5e-80;
Matches 152; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MVTPTWMDNPIEVYVNDVSVVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60
Db 1 MVTPTWMDNPIEVYVNDVSVVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60
Qy 61 LMPAVQNNLVVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPKREI 120
Db 61 LMPAVQNNLVVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPKREI 120
Qy 121 PGSKNTEVLWEECVANSVVLQNNFEFTIID 153
Db 121 PGSKNTEVLWEECVANSVVLQNNFEFTIID 153
RESULT 4
O9UNW3
ID O9UNW3 PRELIMINARY; PRT; 153 AA.
AC O9UNW3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE ENVELOPE PROTEIN RIC-2.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449692; PubMed=9778243;
RA Lan M.S., Mason A., Coutant R., Chen Q.Y., Vargas A., Rao J.,
Gomez R., Chalew S., Garry R., McLaren N.K.;
RT "HERV-K10s and immune-mediated (type 1) diabetes.";
RL Cell 95:14-16(1998).
DR EMBL; AF084865; AAC68894.1; -;
KW Envelope protein.
SQ SEQUENCE 153 AA; 17226 MW; 4632F483BFD9517A CRC64;

Query Match 96.5%; Score 811; DB 4; Length 153;
Best Local Similarity 96.7%; Pred. No. 1.7e-78;
Matches 148; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Qy 1 MVTPTWMDNPIEVYVNDVSVVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60
Db 1 MVTPTWMDNPIEVYVNDVSVVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60
Qy 61 LMPAVQNNLVVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPKREI 120
Db 61 LMPAVQNNLVVEVPTVSPISRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPKREI 120
Qy 121 PGSKNTEVLWEECVANSVVLQNNFEFTIID 153
Db 121 PGSKNTEVLWEECVANSVVLQNNFEFTIID 153
RESULT 5
O9UNW2
ID O9UNW2 PRELIMINARY; PRT; 153 AA.
AC O9UNW2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE ENVELOPE PROTEIN RIC-7.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```
RP SEQUENCE FROM N.A.
PX MEDLINE=98449692; PubMed=9778243;
RA Lan M.S., Mason A., Coutant R., Chen O.Y., Vargas A., Rao J.,
R Gomez R., Chalew S., Garry R., MacLaren N.K.:
RT "HERV-K10s and immune-mediated (type 1) diabetes.";
RL Cell 95:14-16(1998).
DR EMBL; AF084870; AAC68899.1; -.
KW Envelope protein.
SQ SEQUENCE 153 AA; 17406 MW; 90B91BCE0040C5A3 CRC64;

Query Match 95.8%; Score 805; DB 4; Length 153;
Best Local Similarity 96.7%; Pred. No. 7.2e-78;
Matches 148; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MVTPTWMDNPIEVYVNDVSVVWPGPTDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 60
DB 1 MVTPTWMDNPIEVYVNDVSVVWPGPTDRCAPKPEEGMMINISIVRYPPICLGRAPGC 60
QY 61 LMPAVQNLVLEVPVTPSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120
DB 61 LMPAVQNLVLEVPVTPSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120
QY 121 PKGSKNTEVLWEECVANSVVILQNNFEGTIID 153
DB 121 PKGSKNTEVLWEECVANSVVILQNNFEGTIID 153

RESULT 6
Q9HDB8 PRELIMINARY; PRT; 245 AA.
AC Q9HDB8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ENV.
GN Env.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Jinno Y., Sugimoto J.;
RT "Human endogenous retrovirus HERV-K(II) and flanking sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047240; BAB11760.1; -.
SQ SEQUENCE 245 AA; 27904 MW; 198F26D65ED56DDB CRC64;

Query Match 95.8%; Score 805; DB 4; Length 245;
Best Local Similarity 96.7%; Pred. No. 1.2e-77;
Matches 148; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MVTPTWMDNPIEVYVNDVSVVWPGPTDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 60
DB 1 MVTPTWMDNPIEVYVNDVSVVWPGPTDRCAPKPEEGMMINISIVRYPPICLGRAPGC 60
QY 61 LMPAVQNLVLEVPVTPSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120
DB 61 LMPAVQNLVLEVPVTPSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120
QY 121 PKGSKNTEVLWEECVANSVVILQNNFEGTIID 153
DB 121 PKGSKNTEVLWEECVANSVVILQNNFEGTIID 153

RESULT 7
Q14273 PRELIMINARY; PRT; 1361 AA.
AC Q14273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
```

```
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE POL/ENV ORF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=87036922; PubMed=3021993;
RA Ono M., Yasunaga T., Miyata T., Ushikubo H.;
RT "Nucleotide sequence of human endogenous retrovirus genome related to
RT the mouse mammary tumor virus genome.";
RL J. Virol. 60:589-598(1986).
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
CC TRANSCRIPTASE)
DB EMBL; M14123; AAA88033.1; -.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_zn.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF02022; Integrase_zn; 1.
DR Pfam; PF00075; rnaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1361 AA; 153797 MW; CEB91B3F407B9498 CRC64;

Query Match 94.3%; Score 792; DB 4; Length 1361;
Best Local Similarity 97.8%; Pred. No. 2.1e-75;
Matches 145; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 MVTPTWMDNPIEVYVNDVSVVWPGPTDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 60
DB 774 MVTPTWMDNPIEVYVNDVSVVWPGPTDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 833
QY 61 LMPAVQNLVLEVPVTPSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120
DB 834 LMPAVQNLVLEVPVTPSPISRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 893
QY 121 PKGSKNTEVLWEECVANSVVILQNNFEGTIID 153
DB 894 PKGSKNTEVLWEECVANSVVILQNNFEGTIID 926

RESULT 8
Q9UKH9 PRELIMINARY; PRT; 2294 AA.
AC Q9UKH9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE GAG-PRO-POL-ENV PROTEIN.
DE GAG-PRO-POL-ENV PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99400989; PubMed=10469592;
RA Barbulescu M., Turner G., Seaman M.I., Deinard A.S., Kidd K.K.,
RA Lenz J.;
RT "Many human endogenous retrovirus K (HERV-K) proviruses are unique to
RT humans.";
RL Curr. Biol. 9:861-868(1999).
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
CC TRANSCRIPTASE).
CC -1- SIMILARITY: BELONGS TO 2N-FINGER CCHC TYPE FAMILY.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO THE PEPTIDASE FAMILY A1; ALSO
CC KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.
```

```
DR EMBL; AF164611; AAD51793.1; -.
DR HSP; P16088; IDUT.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001995; Asp_prot_retrov.
DR InterPro; IPR001428; dUTPase.
DR InterPro; IPR003322; Gag_P10.
DR InterPro; IPR000721; Gag_P24.
DR InterPro; IPR000467; G_patch.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_zn.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; Rvise.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00692; dUTPase; 1.
DR Pfam; PF01585; G_patch; 1.
DR Pfam; PF02337; Gag_p10; 1.
DR Pfam; PF00607; gag_p24; 1.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF02022; Integrase_zn; 1.
DR Pfam; PF00075; rnaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; rvp; 1.
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF00098; zf_CCHC; 2.
DR ProDom; PD000946; dUTPase; 1.
DR SMART; SM00443; G_patch; 1.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS00175; ASP_PROT_RETROV; 1.
DR Aspartyl protease; Hydrolase; RNA-directed DNA polymerase;
KW Zinc-finger.
KW SEQUENCE 2294 AA; 257099 MW; 425C81FDDF8A80D0 CRC64;

Query Match 93.9%; Score 789; DB 4; Length 2294;
Best Local Similarity 94.8%; Pred. No. 7.9e-75;
Matches 145; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 MVTPTWMDNPIEVYVNDVSVVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60
Db 1795 MVTPTWMDNPIEVYVNDVSVVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 1854

Qy 61 LMPAVQNMLVEVPTVSPNSRRTYHVMVSGMSLRPRVNYLQDFSQYRSLKFRPKGKTCPKKEI 120
Db 1855 LMPAVQNMLVEVPTVSPISRTYHVMVSGMSLRPRVNYLQDFSQYRSLKFRPKGKTCPKKEI 1914

Qy 121 PGSKNTEVLWEECVANSVILQNNRFGTTID 153
Db 1915 PRESKNTVLWEECVDNSAVILQNNRFGTTID 1947

RESULT 9
O95284
AC O95284 PRELIMINARY; PRT; 153 AA.
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ENVELOPE PROTEIN RIC-6.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449692; PubMed=9778243;
RA Lan M.S., Mason A., Coutant R., Chen Q.Y., Vargas A., Rao J.,
RA Gomez R., Chalew S., Garry R., Maclaren N.K.;
RT "HERV-K10s and immune-mediated (type 1) diabetes.";
RL Cell 95:14-16(1998).
DR EMBL; AF084869; AAC68898.1; -.
DR Envelope protein.
KW SEQUENCE 153 AA; 17360 MW; 28C01AD348C6D806 CRC64;

Query Match 93.1%; Score 782; DB 4; Length 153;
Best Local Similarity 94.1%; Pred. No. 2e-75;
Matches 144; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MVTPTWMDNPIEVYVNDVSVVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60
Db 1 MVTPTWMDNPIEVYVNDSENVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGTAPGC 60

Qy 61 LMPAVQNMLVEVPTVSPNSRRTYHVMVSGMSLRPRVNYLQDFSQYRSLKFRPKGKTCPKKEI 120
Db 61 LMPAVQNMLVEVPTVSPISRTYHVMVSGMSLRPRVNYLQDFSQYRSLKFRPKGKTCPKKEI 120

Qy 121 PGSKNTEVLWEECVANSVILQNNRFGTTID 153
Db 121 PRESKNTVLWEECVANSVILQNNRFGTTID 153

RESULT 11
O9YNA9
ID O9YNA9 PRELIMINARY; PRT; 694 AA.
AC O9YNA9;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ENV PROTEIN.
GN ENV.
OS Human endogenous retrovirus K.
```


OC Viruses; Retroid viruses; Retroviridae.
 OX NCBI_TaxID=45617;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Toenjes R.R., Crauderna F., Kurth R.;
 RT "Full-length human endogenous retrovirus type K (HERV-K) elements
 RT encoding Gag, Pol and Env proteins are localised on chromosomes 7 and
 RT 19.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Y17832; CAA76880.1; -;
 DR EMBL: Y17832; CAA76880.1; -;
 SQ SEQUENCE 694 AA; 78690 MW; 82F91825669CF25B CRC64;

Query Match 92.7%; Score 779; DB 12; Length 694;
 Best Local Similarity 92.8%; Pred. No. 2.3e-74;
 Matches 142; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MVTPTWMDNPIEVYVNDVSVVPGTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60
 Db 107 LIRAVTWMNDNTEVYVNDVSVVPGTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 166
 Qy 61 LMPAVQNLWLVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPCKEI 120
 Db 167 LMPAVQNLWLVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPCKEI 226
 Qy 121 PKGSKNTEVLWEECVANSVILQNNFEFTIID 153
 Db 227 PKGSKNTEVLWEECVANSVILQNNFEFTIID 259

RESULT 12
 Q9UBU4
 ID Q9UBU4 PRELIMINARY; PRT; 699 AA.
 AC Q9UBU4;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE ENVELOPE PROTEIN.
 GN ENV.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99400989; PubMed=10469592;
 RA Barbulesscu M., Turner G., Seaman M.I., Deinard A.S., Kidd K.K.,
 RA Lenz J.;
 RT "Many human endogenous retrovirus K (HERV-K) proviruses are unique to
 RT humans.";
 RL Curr. Biol. 9:861-868(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99178257; PubMed=10080172;
 RA Mayer J., Sauter M., Racz A., Scherer D., Mueller-Lantzsch N.,
 RA Meese E.;
 RT "An almost-intact human endogenous retrovirus K on human chromosome
 RT 7.";
 RL Nat. Genet. 21:257-258(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Reus K., Mayer J., Sauter M., Scherer D., Mueller-Lantzsch N.,
 RA Meese E.;
 RT "Further characterization of the almost intact human endogenous
 RT retrovirus K on human chromosome 7.";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF164614; AA051798.1; -;
 DR EMBL: AF164614; AA051798.1; -;
 DR EMBL: AF074086; AAF88168.1; -;
 DR EMBL: AF074086; AAD21098.1; -;
 KW Envelope protein.
 SQ SEQUENCE 699 AA; 79217 MW; 3844DA6D80D76DFC CRC64;

Query Match 92.7%; Score 779; DB 4; Length 699;
 Best Local Similarity 92.8%; Pred. No. 2.3e-74;
 Matches 142; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
 Qy 1 MVTPTWMDNPIEVYVNDVSVVPGTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60
 Db 112 LIRAVTWMNDNTEVYVNDVSVVPGTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 171
 Qy 61 LMPAVQNLWLVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPCKEI 120
 Db 172 LMPAVQNLWLVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPCKEI 231
 Qy 121 PKGSKNTEVLWEECVANSVILQNNFEFTIID 153
 Db 232 PKGSKNTEVLWEECVANSVILQNNFEFTIID 264

RESULT 13
 Q69384
 ID Q69384 PRELIMINARY; PRT; 699 AA.
 AC Q69384;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE ENV MRNA.
 GN ENV.
 OS Human endogenous retrovirus.
 OC Viruses; Retroid viruses; Retroviridae.
 OX NCBI_TaxID=11827;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95074858; PubMed=7983704;
 RA Loewer R., Toenjes R.R., Korbacher C., Kurth R., Loewer J.;
 RT "Identification of a Rev-related protein by analysis of spliced
 RT transcripts of the human endogenous retroviruses HTDV/HERV-K.";
 RL J. Virol. 69:141-149(1995).
 DR EMBL: X82272; CAA57723.1; -;
 SQ SEQUENCE 699 AA; 79217 MW; 3844DA6D80D76DFC CRC64;

Query Match 92.7%; Score 779; DB 12; Length 699;
 Best Local Similarity 92.8%; Pred. No. 2.3e-74;
 Matches 142; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
 Qy 1 MVTPTWMDNPIEVYVNDVSVVPGTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60
 Db 112 LIRAVTWMNDNTEVYVNDVSVVPGTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 171
 Qy 61 LMPAVQNLWLVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPCKEI 120
 Db 172 LMPAVQNLWLVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPCKEI 231
 Qy 121 PKGSKNTEVLWEECVANSVILQNNFEFTIID 153
 Db 232 PKGSKNTEVLWEECVANSVILQNNFEFTIID 264

RESULT 14
 Q69386
 ID Q69386 PRELIMINARY; PRT; 514 AA.
 AC Q69386;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE POL/ENV GENE (FRAGMENT).
 GN POL/ENV.
 OS Human endogenous retrovirus K.
 OC Viruses; Retroid viruses; Retroviridae.
 OX NCBI_TaxID=45617;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97213939; PubMed=9060628;
 RA Toenjes R.R., Limbach C., Lower R., Kurth R.;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 16:54:06 ; Search time 39.41 Seconds
(without alignments)
295.730 Million cell updates/sec

Title: US-09-490-700-36
Perfect score: 840
Sequence: 1 MVPTVWMDNPIEVYVNDV.....ECVANSVVILQNNFGTTID 153

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	768.5	91.5	584	4 VCHUER	retrovirus-related
2	82	9.8	226	2 A53273	MHC class II histo
3	81.5	9.7	990	2 T02309	probable transcrip
4	76	9.0	492	2 T26502	hypothetical prote
5	76	9.0	514	2 T26501	hypothetical prote
6	75.5	9.0	896	1 A35782	cytokine receptor
7	74.5	8.9	318	2 S76295	hypothetical prote
8	74	8.8	602	2 S39782	cyclooxygenase 1 -
9	74	8.8	466	2 A36674	transcription fact
10	73.5	8.8	297	2 T36724	probable membrane
11	73	8.7	1072	2 S76888	hypothetical prote
12	72.5	8.6	365	2 C86647	hypothetical prote
13	72.5	8.6	602	2 S69198	prostaglandin G/H
14	72.5	8.6	1475	2 A44765	alpha-amyrase (EC
15	72	8.6	341	2 T45350	hypothetical prote
16	72	8.6	688	1 VCMVM	env polyprotein -
17	72	8.6	848	2 C65083	hypothetical prote
18	71.5	8.5	481	2 S04605	glycinin G3 - soyb
19	71.5	8.5	484	2 S11003	glycinin G3 precu
20	71.5	8.5	1003	1 PVV2AM	spheroidin precurs
21	71.5	8.5	1190	2 S21977	Fm5 protein - huma
22	71.5	8.5	1262	2 T25168	hypothetical prote
23	71	8.5	251	2 B83413	molybdate-binding
24	70.5	8.4	295	2 F83201	conserved hypothet
25	70.5	8.4	387	2 B55164	scn1 protein - fis
26	70	8.3	256	2 T51150	probable transmemb
27	70	8.3	491	2 B86155	env polyprotein -
28	70	8.3	688	2 S26388	malate synthase (E
29	69.5	8.3	555	1 SYHOMA	

RESULT 1
VCHUER
retrovirus-related env polyprotein pseudogene - human
C:Species: Homo sapiens (man)
C:Date: 28-Dec-1987 #sequence_revision 04-Jan-1996 #text_change 14-May-1999
C:Accession: E24483
R:Ono, M.; Yasunaga, T.; Miyata, T.; Ushikubo, H.
J. Virol. 60, 589-598, 1986
A:Title: Nucleotide sequence of human endogenous retrovirus genome related to the mou
A:Reference number: A93023; MUID:87036922
A:Accession: E24483
A:Status: conceptual translation of pseudogene
A:Molecule type: DNA
A:Residues: 1-584 <ONO>
A:Cross-references: GB:M14123; NID:g182227
C:Genetics:
A:Gene: env
C:Keywords: capsid protein; coat protein; polyprotein; pseudogene

ALIGNMENTS

30	69.5	8.3	623	2	H96766	unknown protein F2
31	69.5	8.3	1274	2	T16251	hypothetical prote
32	69.5	8.3	1481	2	S28669	anylopolulanase p
33	69	8.2	204	2	T02386	hypothetical prote
34	69	8.2	1520	2	T00273	hypothetical prote
35	68.5	8.2	549	2	H86024	cytoplasmic trehal
36	68.5	8.2	549	2	S47739	probable alpha,alp
37	68	8.1	214	1	RGBSCA	regulatory protein
38	68	8.1	355	2	T29659	hypothetical prote
39	68	8.1	688	1	VCMVM	env polyprotein -
40	68	8.1	876	2	PC2219	polypeptide - hepa
41	68	8.1	1210	2	S35548	DNA-directed RNA p
42	67.5	8.0	170	2	B71510	probable methyltra
43	67.5	8.0	285	2	T29490	hypothetical prote
44	67.5	8.0	473	2	S70357	forkhead transcrip
45	67.5	8.0	479	2	T01922	hypothetical prote

Query Match 91.5%; Score 768.5; DB 4; Length 584;
Best Local Similarity 94.7%; Pred. No. 4.7e-69;
Matches 142; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

Qy 4 PVTWMDNPIEVYVNDVWVPGTDDRCAPKEEGGMINISIGYHYPPICLGRAPGCLMP 63
Db 1 PVTWMDNPIEVYVNDVWVPGTDDRCAPKEEGGMINISIGYHYPPICLGRAPGCLMP 60
Qy 64 AVQNLVEVPTVSPNSRFTYHMVSGMSLRPRVNLQDFSYQSLKFRPKGTCPEIPKG 123
Db 61 AVQNLVEVPTVSPISRFTYHMVSGMSLRPRVNLQDFSYQSLKFRPKGTCPEIPKE 120
Qy 124 SKNTEVLWEECVANSVVILQNNFGTTID 153
Db 121 SKNTEVLWEECVANSVIL-NNFGTTID 149

RESULT 2
A53273
MHC class II histocompatibility antigen DR alpha chain - horse (fragment)
C:Species: Equus caballus (domestic horse)
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 21-Jan-2000
C:Accession: A53273
R:Albright, D.; Bailey, E.; Woodward, J.G.
Immunogenetics 34, 136-138, 1991
A:Title: Nucleotide sequence of a cDNA clone of the horse (Equus caballus) DRA gene.
A:Reference number: A53273; MUID:91331619
A:Accession: A53273
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-226 <ALB>
A:Cross-references: GB:M60100; NID:g164236; PIDN:AAA30956.1; PID:g164237

C;Genetics:

A;Gene: DRA
A;Map position: 20
C;Superfamily: class II histocompatibility antigen; immunoglobulin homology
F;97-162/Domain: immunoglobulin homology <IMW>

Query Match 9.8%; Score 82; DB 2; Length 226;
Best Local Similarity 27.7%; Pred. No. 0.81;
Matches 28; Conservative 16; Mismatches 29; Indels 28; Gaps 5;

Qy 2 VPTVMDN--PIEYVNDVSVVPGPTDDRCAPKPEEGMMINISIGYHYHPPPI----- 52

Db 113 VIDVTWLNKGRPVMTGVSETVFL--PRDQLFRK-----FHYLPFLPSTEDV 157

Qy 53 --CLGRAPCLMPAVONWLVVPTVSPNSRFTYHMVSCMSL 91

Db 158 YDCKVEHWGLDEPLKHWFEAPT--PUSETTENNVCGGL 196

RESULT 3

T02309
probable transcription elongation factor [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein F13P17.5
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 02-Mar-2001
C;Accession: T02309; G84753
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, July 1998
A;Description: Arabidopsis thaliana chromosome II BAC F13P17 genomic sequence.
A;Reference number: Z14657
A;Accession: T02309
A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA
A;Residues: 1-990 <ROU>
A;Cross-references: EMBL:AC004481; NID:g3337347; PID:g3337352
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487

A;Accession: G84753
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-990 <STO>
A;Cross-references: GB:AE002093; NID:g3337352; PIDN:AAC27397.1; GSPDB:GN00139
C;Genetics:
A;Gene: F13P17.5; At2g34210
A;Map position: 2
A;Introns: 97/1; 177/3; 226/3; 280/3; 308/3; 332/2; 451/3; 503/3; 536/2; 552/3; 603/3; 6
C;Keywords: transcription factor

Query Match 9.7%; Score 81.5; DB 2; Length 990;
Best Local Similarity 21.1%; Pred. No. 5.1;
Matches 37; Conservative 34; Mismatches 65; Indels 39; Gaps 7;

Qy 4 PVTWMDNPIEYVNDVSVV---VPGPTDDRCAPKPEEGMMINISIGYHYHPPICL-GRAPG 59

Db 806 PGSGWGTSPYEATPGSDWGSSTPGRSSYRDAGTP-----INNGFYVYLLCLNANAPS 857

Qy 60 CLMPAVONWLVVPTVSPNSRFTYHMVSCMSLPRVNYL-----QDFSQRSL 107

Db 858 PWTFSSTSYLPTPGCGQAMTGTDLVMSLDIVNIFQTDIVSUFFLCGHQDGSCVVAL 917

Qy 108 KFRPKGKT-----CPKEIPKSKNTEVLVWEECVANSVVLQNNFEFTI 151

Db 918 GHRGEGETIRATQNKVSLVCPKK-----NERVKILGGKYCGSTAKVIGEDGGDGI 968

RESULT 4

T26502

hypothetical protein Y17G7B.10b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26502
R;Smye, R.
submitted to the EMBL Data Library, June 1998
A;Reference number: Z20225
A;Accession: T26502
A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-492 <WIL>

A;Cross-references: EMBL:AL023828; PIDN:CAA19456.1; GSPDB:GN00020; CESP:Y17G7B.10b

A;Experimental source: clone Y17G7B

C;Genetics:

A;Gene: CESP:Y17G7B.10b

A;Map position: 2

A;Introns: 12/3; 60/1; 124/3; 154/2; 240/2; 298/2; 423/1

Query Match 9.0%; Score 76; DB 2; Length 492;
Best Local Similarity 26.0%; Pred. No. 8;
Matches 40; Conservative 21; Mismatches 41; Indels 52; Gaps 10;

Qy 4 PVT--WMDNPIEYVNDVSVVPGPTDDRCAPKPEEGMMINISIGYHYHPPI---CLGRAP 58

Db 272 PATTRWTDNP-----RSPAREIEECNLSL-----PPVKRKRMTSEA- 306

Qy 59 GCLMPAVONWLVVPTVSPNSRFTYHMVSCMSLPRVNYLQ-DFSYSRSLKFRPKGKTCP 117

Db 307 -VEIPEVEKETVELPVLALNEVFI-----GESLSRSVSYEIGINDAQMALKOKSSGIT-- 358

Qy 118 KEIPKSKNTEVLVW-----EECVANSVVIL 143

Db 359 --ICTGTGSTS---WNFNINKLTEQCVDLMKIV 387

RESULT 5

T26501
hypothetical protein Y17G7B.10a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26501
R;Smye, R.
submitted to the EMBL Data Library, June 1998
A;Reference number: Z20225
A;Accession: T26501
A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-514 <WIL>

A;Cross-references: EMBL:AL023828; PIDN:CAA19455.1; GSPDB:GN00020; CESP:Y17G7B.10a

A;Experimental source: clone Y17G7B

C;Genetics:

A;Gene: CESP:Y17G7B.10a

A;Map position: 2

A;Introns: 15/2; 34/3; 82/1; 146/3; 176/2; 262/2; 320/2; 445/1

Query Match 9.0%; Score 76; DB 2; Length 514;
Best Local Similarity 26.0%; Pred. No. 8.4;
Matches 40; Conservative 21; Mismatches 41; Indels 52; Gaps 10;

Qy 4 PVT--WMDNPIEYVNDVSVVPGPTDDRCAPKPEEGMMINISIGYHYHPPI---CLGRAP 58

Db 294 PATTRWTDNP-----RSPAREIEECNLSL-----PPVKRKRMTSEA- 328

Qy 59 GCLMPAVONWLVVPTVSPNSRFTYHMVSCMSLPRVNYLQ-DFSYSRSLKFRPKGKTCP 117

Db 329 -VEIPEVEKETVELPVLALNEVFI-----GESLSRSVSYEIGINDAQMALKOKSSGIT-- 380

Qy 118 KEIPKSKNTEVLVW-----EECVANSVVIL 143

Db 381 --ICTGTGSTS---WNFNINKLTEQCVDLMKIV 409

```

Qy      119 EIPKGSK 125
      | | |
Db      311 ESPGKE 317

RESULT      8
S39782
Cyclooxygenase 1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-May-1994 #sequence_revision 10-Nov-1995 #text_change 10-Dec-1999
C:Accession: S39782
R:Feng, L.; Sun, W.; Xia, Y.; Tang, W.W.; Channmugam, P.; Soyoola, E.; Wilson, C.B.; H
A:Title: Cloning two isoforms of rat cyclooxygenase: differential regulation of their
A:Reference number: S39782; MUID:94099619
A:Accession: S39782
A:Molecule type: mRNA
A:Residues: 1-602 <FEN>
A:Cross-references: GB:S67721; NID:9460355; PIDN:AAB29400.1; PID:9460356
C:Superfamily: human prostaglandin-endoperoxide synthase; EGF homology
F:38-71/Domain: EGF homology <EGF>

```

Query Match	8.9%;	Score 74.5;	DB 2;	Length 602;
Best Local Similarity	23.3%.	pred No. 14.		

Qy 11 PIEVYVNDVWVPGPTDDRCPAKPEEGMMINTISGYHYPPICLG---RAPGCLMPAVQN 67

db 21 PPPVLLTDA-GVPSPVIPCCYYPCQNQGVCVREFGLD-HYOCDCTRTGYSGPNCTIPEIWT 78

Qy 68 WLVEVP TVSP-----NSRFTYH MVSGMSLRPRVNYL----- 98

Db 79 WLRSSLRPSSETHELLTHGYWIEFVNATFIREVLMGWVLTVRSNLTPSPPTYNTAHDY 138

COMMITTEE TO REFORM THE HOUSE OF REPRESENTATIVES

QY 99 ---QDFS---YQPSLFRPVGKTCPEIPKSGSKNTEVL 130

```

Db      139 ISWESFSNVSYTRIIPSPV--KDCP--TPMGTKGKKOI. 173

```

RESULT 9

RESOLVED
A36674

transcription factor HNF-3A, hepatic - rat

N; Alternate names: nuclear factor 3A, hepatic

C:species: rattus norvegicus (NOIRWAY IAL)
C:Date: 12-Apr-1991 #sequence revision 12-Apr-1991 #text change 16-Jul-1999

C;Accession: A36674; S15651

R; Lai, E.; Prezioso, V.R.; Smith, E.; Litvin, O.; Costa, R.H.; Darnell Jr., J.E.
Genes Dev 4 1427-1436 1990

Genes Dev. 4, 1427-1430, 1990

A;Reference number: A36674; MUID:91032994

A;Accession: A36674
A.Molecule type: mRNA

A; Molecule type: MINNA
A; Residues: 1-466 <LAI>

A;Cross-references: GB:X55955; NID:g56369; PIDN:CAA39418.1; PID:g56370

C; Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology

C;keyworus: DNA binding; liver; nucleus; transcription factor

F;170-261/Domain: fork head DNA-binding domain homology <FHD>

Query Match	8.8%	Score 74;	DB 2;	Length 466;
Best Local Similarity	25.2%	Pred. No. 12;		

Matches 29; Conservative 13; Mismatches 57; Indels 16; Gaps 3;

0v 23 pcr-----TDDRCBPAKDEECMMINISICVHVDDICICBAOCCIMDAVONTI VE-----71

QJ	ZS	FOF	IDB	KCF	RNF	ELE	GNT	MINS	IGI	HIF	FIC	BNA	PCE	MPA	VQN	WLV	E - - - - - / I
									:								

Db 322 PCPAASPQTLDHSGATATCGGSELKSPASSAPPISSGPGGWICTPLSPTWLAPHESQLH 381

73 - - - - - VOTEC/CONCERNED BY UNLAWFUL CONDUCT OF PERSONS WHOSE NAMES ARE LISTED IN THE

QY 72 --VFIVSPNSRFTTHMVSGMSLRPRVNYLQDFSIQRSLRFRPRGKTCPRKPG 123

Db 382 LKGAPHYSFNHPFSINNL--MSSSEQQHKLDFKAYEQALQYSPYGATLPASPLG 434

```

Query Match      8.8%; Score 73.5; DB 2; Length 297;
Best Local Similarity 33.9%; Pred. NO. 7.9;
Matches 20; Conservative 6; Mismatches 18; Indels 15; Gaps 3;

Qy 18 DSVWPG-----PTDRC-----PAKPEEGMINISGYHYPPICLGRAPGCLMPA 64
   : :||| ||| | | | : | : | : | | | | | | | |
Db 10 DAQSLPGCYRHPDRETGIRCTRCERPCPD---CMVNASVGFHCPDCAAGRTFGATGPA 65

RESULT 11
S76888

```

C, Species: *Synechocystis* sp.
C, Variety: PCC 6803
C, Date: 25-Apr-1997 #sequence_revision 25-May-1997 #text_change 20-Jun-2000
C, Accession: S76888
R; Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, K.; Tsumoto, K. J. Mol. Biol. 267: 125-132, 1998
DNA Res. 3, 109-136, 1996
A, Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.
A, Reference number: S74322; MUID:97061201
A, Accession: S76888
A, Status: preliminary
A, Molecule type: DNA
A, Residues: 1-1072 <KAN>
A, Cross-references: EMBL:D90917; GB:AB001339; NID:g1653836; PIDN:BAAL18800.1; PID:g1653836
A, Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C, Superfamily: *Synechocystis* hypothetical protein sl10446

Query Match 8.7%; Score 73; DB 2; Length 1072;
Best Local Similarity 28.2%; Pred. No. 40;
Matches 42; Conservative 16; Mismatches 71; Indels 20; Gaps 8;

Qy 1 MVTPTVMNDNP -IEVYND----SVWVPGPTDDRCPAK--PEEGMMIN-ISTGYHVPPI 52
||| | | | : : : | | | | | | : | |
Db 130 MVPP-----DGPWAELYFDLGVAGMTSPATIVCPGTGYFPEHLSSRIAWIKNGFFTNPI 185

Db 186 GNGIAPNKOKEVLLPWLNLKGNLMR-NPQNEDLAGRVSQMI-----DDYIQDLGVTTSNVF 240
Qy 110 RPKGKTCPEIKPGSKNTEVLVWEECVAN 138
Db 241 QPTTKNLFFGIPLAPPPLNALYPAEVQN 269

RESULT 12
C86647
hypothetical protein yb1d [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis

Search completed: April 9, 2002, 16:58:10
Job time: 244 sec

```

Query Match      8 6%; Score 72; DB 2; Length 341;
Best Local Similarity 25.3%; Pred. NO. 13;
Matches 24; Conservative 13; Mismatches 37; Indels 20; Gaps 4;

QY 3 TPVTWMDNPTEVYXVNDVWVGGPTD---DRCP---AKPEEGCMINI-----SIGYHYP 50
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 42 TDIDESPFAVANDNPRTWLTPTDPLGRHPWVLAQPEQRKIEITGMVQANVAVKVLHFE 101

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 16:53:21 ; Search time 34.73 Seconds
(without alignments)
99.136 Million cell updates/sec

Title: US-09-490-700-36

Perfect score: 840
Sequence: 1 MVTPTWMDNPIEVVDSV.....ECVANSVILQNEFTIID 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71.5	8.5	1003	1 US-07-991-867B-6	Sequence 6, Appl
2	71.5	8.5	1003	1 US-08-107-755A-6	Sequence 6, Appl
3	71.5	8.5	1003	2 US-08-544-332-6	Sequence 6, Appl
4	67.5	8.0	473	4 US-08-857-076-99	Sequence 99, Appl
5	67.5	8.0	484	3 US-09-080-044-7	Sequence 7, Appl
6	67.5	8.0	533	1 US-08-220-151-15	Sequence 15, Appl
7	67.5	8.0	533	1 US-08-413-118-15	Sequence 15, Appl
8	67.5	8.0	533	3 US-08-473-446-15	Sequence 15, Appl
9	66.5	7.9	480	4 US-09-078-173A-25	Sequence 15, Appl
10	66.5	7.9	480	4 US-09-537-357-32	Sequence 32, Appl
11	66.5	7.9	807	1 US-07-862-021B-10	Sequence 10, Appl
12	66.5	7.9	807	1 US-08-313-288B-10	Sequence 10, Appl
13	66.5	7.9	807	5 PCT-US93-03164-10	Sequence 10, Appl
14	66.5	7.9	2409	6 5180808-2	Patent No. 5180808
15	65.5	7.8	410	4 US-09-258-754-451	Sequence 451, App
16	65	7.7	338	1 US-08-218-686-2	Sequence 2, Appl
17	65	7.7	338	3 US-08-460-242-2	Sequence 2, Appl
18	64.5	7.7	802	1 US-07-862-021B-12	Sequence 12, Appl
19	64.5	7.7	802	1 US-08-313-288B-12	Sequence 12, Appl
20	64.5	7.7	802	5 PCT-US93-03164-12	Sequence 12, Appl
21	64	7.6	509	3 US-08-688-988-40	Sequence 40, Appl
22	64	7.6	1420	2 US-08-540-804-14	Sequence 14, Appl
23	64	7.6	1420	2 US-08-218-265-14	Sequence 14, Appl
24	64	7.6	1420	3 US-08-521-872-14	Sequence 14, Appl
25	64	7.6	1420	4 US-08-590-399-14	Sequence 14, Appl
26	63.5	7.6	326	3 US-08-808-720-3	Sequence 3, Appl
27	63.5	7.6	328	3 US-08-808-720-1	Sequence 1, Appl

28	63.5	7.6	331	3 US-08-808-720-5	Sequence 5, Appl
29	63.5	7.6	331	3 US-08-808-720-7	Sequence 7, Appl
30	63.5	7.6	467	4 US-08-523-894-10	Sequence 10, Appl
31	63	7.5	192	1 US-08-086-428B-87	Sequence 87, Appl
32	63	7.5	192	2 US-08-468-570-87	Sequence 87, Appl
33	63	7.5	192	2 US-08-290-665A-87	Sequence 87, Appl
34	63	7.5	192	5 PCT-US95-10398-87	Sequence 14, Appl
35	63	7.5	516	4 US-08-868-373-14	Sequence 26, Appl
36	62.5	7.4	128	2 US-08-470-139-26	Sequence 27, Appl
37	62.5	7.4	248	1 US-08-644-664B-27	Sequence 27, Appl
38	62.5	7.4	248	2 US-08-761-277A-27	Sequence 27, Appl
39	62.5	7.4	382	1 US-08-470-299-10	Sequence 10, Appl
40	62.5	7.4	467	4 US-08-523-894-12	Sequence 12, Appl
41	62.5	7.4	1182	4 US-09-041-886-21	Sequence 21, Appl
42	62	7.4	192	1 US-08-086-428B-100	Sequence 100, App
43	62	7.4	192	2 US-08-468-570-100	Sequence 100, App
44	62	7.4	192	2 US-08-290-665A-100	Sequence 100, App
45	62	7.4	192	5 PCT-US95-10398-100	Sequence 100, App

ALIGNMENTS

RESULT 1
US-07-991-867B-6
; Sequence 6, Application US/07991867B
; Patent No. 5476781
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/991,867B
; FILING DATE: 12-DEC-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFL14.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-991-867B-6

Query Match 8.5%; Score 71.5; DB 1; Length 1003;
Best Local Similarity 25.0%; Pred. No. 11;
Matches 26; Conservative 12; Mismatches 37; Indels 29; Gaps 5;

```

QY 60 CLMPAVQ-----NWLEVEPTSPNSRFFYHMVSG---MSLRPRVNYLQDFSVQSLKPR 110
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 CLKPKPKNLRLNGWLLDCT-----SRFTKHMADGSDDLQDLVRLN----- 482
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 111 PKGKTCPEK-IPGSKNTEVLVWEEGVANSVVILQNEEFTIID 153
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 483 -RNDICLKQAIKHGVYNVILLEYANTYPNCTSLGNRRNNPNF 525
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

1  RESULT      2
2  US-08-107-755A-6
3  ; Sequence 6, Application US/08107755A
4  ; Patent No. 5721352
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Moyer, Richard W.
7  ; APPLICANT: Hall, Richard L.
8  ; APPLICANT: Gruidl, Michael E.
9  ; TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
10 ; NUMBER OF SEQUENCES: 40
11 ; CORRESPONDENCE ADDRESS:
12 ; ADDRESSEE: David R. Saliwanchik
13 ; STREET: 2421 N.W. 41st Street, Suite A-1
14 ; CITY: Gainesville
15 ; STATE: Florida
16 ; COUNTRY: U.S.A.
17 ; ZIP: 32606
18 ; COMPUTER READABLE FORM:
19 ; MEDIUM TYPE: Floppy disk
20 ; COMPUTER: IBM PC compatible
21 ; OPERATING SYSTEM: PC-DOS/MS-DOS
22 ; SOFTWARE: Patent In Release #1.0, Version #1.25
23 ; CURRENT APPLICATION DATA:
24 ; APPLICATION NUMBER: US/08/107,755A
25 ; FILING DATE: 19-AUG-1993
26 ; CLASSIFICATION: 435
27 ; PRIOR APPLICATION DATA:
28 ; APPLICATION NUMBER: US 07/827,658
29 ; FILING DATE: 30-JAN-1992
30 ; PRIOR APPLICATION DATA:
31 ; APPLICATION NUMBER: US 07/657,584
32 ; FILING DATE: 19-FEB-1991
33 ; ATTORNEY/AGENT INFORMATION:
34 ; NAME: Saliwanchik, David R.
35 ; REGISTRATION NUMBER: 31,794
36 ; REFERENCE/DOCKET NUMBER: UF114.C2
37 ; TELECOMMUNICATION INFORMATION:
38 ; TELEPHONE: (904) 375-8100
39 ; TELEFAX: (904) 372-5800
40 ; INFORMATION FOR SEQ ID NO: 6:
41 ; SEQUENCE CHARACTERISTICS:
42 ; LENGTH: 1003 amino acids
43 ; TYPE: amino acid
44 ; TOPOLOGY: linear
45 ; MOLECULE TYPE: protein
46 US-08-107-755A-6

```

Query Match 8.5%; Score 71.5; DB 1; Length 1003;
Best Local Similarity 25.0%; Pred. No. 11;
Matches 26; Conservative 12; Mismatches 37; Indels 29; Gaps 5;

```

Qy 60 CLMPAVO-----NWLVVEVTPVSPNGRFTYHKVSG--MSLRPRVNYLODFSQVSLKPR 110
      |||      |||      |||      |||      |||      |||      |||      |||
Db 441 CLPKPKVKNLRLGWLDCDT-----SRFIKHMADGSDDLLOLVKLN----- 482
      |||      |||      |||      |||      |||      |||      |||      |||

Qy 111 PKGTCPEK-IPGSKNTEVLVWEECVANSVVIQLONNEFTIID 153
      |||      |||      |||      |||      |||      |||      |||      |||
Db 483 -RNDICLKQAIKQHYTNVILEYANTVPNCTLSLGNRNENNVED 525
      |||      |||      |||      |||      |||      |||      |||      |||

```

```

RESULT      3
US-08-544-332-6
: Sequence 6, Application US/08544332
: Patent No. 5935777
: GENERAL INFORMATION:
: APPLICANT: Moyer, Richard W.
: APPLICANT: Hall, Richard L.
: APPLICANT: Gruidl, Michael E.
: TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
: NUMBER OF SEQUENCES: 77
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Gerard H. Bencen
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/544,332
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/991,867
: FILING DATE: 07-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/107,755
: FILING DATE: 19-AUG-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO 92/14818
: FILING DATE: 12-FEB-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/827,685
: FILING DATE: 30-JAN-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/657,584
: FILING DATE: 19-FEB-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Bencen, Gerard H.
: REGISTRATION NUMBER: 35,746
: REFERENCE/DOCKET NUMBER: UFI14.C4
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1003 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PS-08-544-332-6

```

Query Match 8.5%; Score 71.5; DB 2; Length 1003;
Best Local Similarity 25.0%; Pred. No. 11;
Matches 26; Conservative 12; Mismatches 37; Indels 29; Gaps 5;

[illegible]

RESULT 4

QY	48	HYPPICLGR-----APGCLMPAVONWLVVEPTVSPNSRFT-----	82
Db	327	HESPVSYKRFKTSVAPDVYP-----PTSVTFADTRAICDVKCVPRDGISLMWKI	377
QY	83	--YHMVSGMSL-----RPRVNYLQ---DFSYORSKLFKPKGKTC-----PKKIPK	122
Db	378	GNVHLPRKMSADILITGPCIERPGLVNIQSMCDISETDG-----PVSYTQITGYPPILP-	432
QY	123	GSKNTEV-----LVWEECVANSVIL 143	
Db	433	GFYDTQVYDASPEIVSESLVSVAVI 459	
<p>RESULT 6</p> <p>US-08-220-151-15</p> <p>; Sequence 15, Application US/08220151</p> <p>; Patent No. 5525780</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Paoletti, Enzo</p> <p>; APPLICANT: Limbach, Keith J.</p> <p>; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF</p> <p>; TITLE OF INVENTION: CANINE HERPESVIRUS gB, gC AND gD AND USES THEREFOR</p> <p>; NUMBER OF SEQUENCES: 91</p> <p>; CORRESPONDENCE ADDRESS:</p> <p>; ADDRESSEE: Curtis, Morris & Safford</p> <p>; STREET: 530 Fifth Avenue</p> <p>; CITY: New York</p> <p>; STATE: NY</p> <p>; COUNTRY: USA</p> <p>; ZIP: 10036</p> <p>; COMPUTER READABLE FORM:</p> <p>; MEDIUM TYPE: Floppy disk</p> <p>; COMPUTER: IBM PC compatible</p> <p>; OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>; SOFTWARE: PatentIn Release #1.0, Version #1.25</p> <p>; CURRENT APPLICATION DATA:</p> <p>; APPLICATION NUMBER: US/08/220,151</p> <p>; FILING DATE: 30-MAR-1994</p> <p>; CLASSIFICATION: 435</p> <p>; ATTORNEY/AGENT INFORMATION:</p> <p>; NAME: Frommer, William S.</p> <p>; REGISTRATION NUMBER: 25,506</p> <p>; REFERENCE/DOCKET NUMBER: 454310-2540</p> <p>; TELECOMMUNICATION INFORMATION:</p> <p>; TELEPHONE: (212) 840-3333</p> <p>; TELEFAX: (212) 840-0712</p> <p>; TELEX: 425066 CURTMS</p> <p>; INFORMATION FOR SEQ ID NO: 15:</p> <p>; SEQUENCE CHARACTERISTICS:</p> <p>; LENGTH: 533 amino acids</p> <p>; TYPE: amino acid</p> <p>; STRANDEDNESS: Single</p> <p>; TOPOLOGY: linear</p> <p>; MOLECULE TYPE: peptide</p> <p>; FRAGMENT TYPE: N-terminal</p> <p>US-08-220-151-15</p>			
<p>Query Match 8.0%; Score 67.5; DB 1; Length 533;</p> <p>Best Local Similarity 23.2%; Pred. No. 13;</p> <p>Matches 48; Conservative 20; Mismatches 54; Indels 85; Gaps</p>			
QY	3	TPVTWM--DNPIE-VYVND--SVWVPG-----PTDRCRCPAKPEEGMMINISIGY	47
Db	321	TKLWYLDGKPIERQYISDTASVWIDGLITRSSVLAIPTTETDSEKPD-----IRCDLEW	375
QY	48	HYPPICLGR-----APGCLMPAVONWLVVEPTVSPNSRFT-----	82
Db	376	HESPVSYKRFKTSVAPDVYP-----PTSVTFADTRAICDVKCVPRDGISLMWKI	426
QY	83	--YHMVSGMSL-----RPRVNYLQ---DFSYORSKLFKPKGKTC-----PKKIPK	122


```
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 807 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-313-288B-10

Query Match          7.9%; Score 66.5; DB 1; Length 807;
Best Local Similarity 23.5%; Pred. No. 32;
Matches 35; Conservative 16; Mismatches 53; Indels 45; Gaps 9;

Qy 14 VYVNDVWVPGPTDDRCAPKEEG-----MMINISIGYHY-----PICLGRAPGC- 60
Db 444 IYSNWSPW-----SACSSSTCEKGRMRQRLKAQLDLSVPCPTQDFQPCMG--PGCS 495
Qy 61 ----LMPAVQNLVEVTPVSPNSRFTYHMYGMSLRPRVNYLQDFSYQSLKFRPKGKT- 115
Db 496 DEDGSTCTMSEWI-----TWSPCS-----VSCGMGRSRERYVKQFPEDGSGVCMPLTETE 546
Qy 116 -----CPKEIPKGSKNTEVLVWEECVA 137
Db 547 KCTVNEEC-----PSSCLVTENGWDCCSA 572

RESULT 13
PCT-US93-03164-10
; Sequence 10, Application PC/TUS9303164
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03164
; FILING DATE: 19930402
```

```
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 807 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-03164-10

Query Match          7.9%; Score 66.5; DB 5; Length 807;
Best Local Similarity 23.5%; Pred. No. 32;
Matches 35; Conservative 16; Mismatches 53; Indels 45; Gaps 9;

Qy 14 VYVNDVWVPGPTDDRCAPKEEG-----MMINISIGYHY-----PICLGRAPGC- 60
Db 444 IYSNWSPW-----SACSSSTCEKGRMRQRLKAQLDLSVPCPTQDFQPCMG--PGCS 495
Qy 61 ----LMPAVQNLVEVTPVSPNSRFTYHMYGMSLRPRVNYLQDFSYQSLKFRPKGKT- 115
Db 496 DEDGSTCTMSEWI-----TWSPCS-----VSCGMGRSRERYVKQFPEDGSGVCMPLTETE 546
Qy 116 -----CPKEIPKGSKNTEVLVWEECVA 137
Db 547 KCTVNEEC-----PSSCLVTENGWDCCSA 572

RESULT 14
5180808-2
; Patent No. 5180808
; APPLICANT: RUOSLAHTI, ERKKI I.
; TITLE OF INVENTION: VERSICAN CORE PROTEIN, NUCLEIC ACID
; SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSICAN
; ANTIBODIES, AND METHODS OF DETECTING THE SAME
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/441,179
; FILING DATE: 27-NOV-1989
; SEQ ID NO:2
; LENGTH: 2409
5180808-2

Query Match          7.9%; Score 66.5; DB 6; Length 2409;
Best Local Similarity 23.0%; Pred. No. 1.5e+02;
Matches 31; Conservative 9; Mismatches 56; Indels 39; Gaps 6;

Qy 7 WMDNPIEVYVNDVWVPGPTDDRCAPKE-----EEMMINISIGYHYHYPPICLGRAPG 59
Db 2252 WTDGSTLTQYEN---WRPNQPDFFSAGEDCVIWHENGQWNDVPCNYHLTYTCKKGTVA 2308
Qy 60 CLMPAVQNLWVE-----VPTVSPNSRFTYHMYGMSLR-----PRVN 96
Db 2309 CGQPPV---VENAKTFGKMKPRYEINSLIRYHCKDGFQIRHLPTIRCLGNRWAIKPT 2364
Qy 97 YLQDFSYOR--SLKF 109
Db 2365 CMNPAYQRTYSMKY 2379

RESULT 15
US-09-258-754-451
```

; Sequence 451, Application US/09258754
; Patent No. 6174687
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; APPLICANT: Rajotte, Daniel
; TITLE OF INVENTION: Methods of Identifying Lung Homing Molecules Using
; FILE OF INVENTION: Membrane Dipeptidase
; FILE REFERENCE: P-LJ 3443
; CURRENT APPLICATION NUMBER: US/09/258,754
; CURRENT FILING DATE: 1999-02-26
; EARLIER APPLICATION NUMBER: 09/042,107
; EARLIER FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 452
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 451
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-258-754-451

Query Match 7.88; Score 65.5; DB 4; Length 410;
Best Local Similarity 25.88; Pred. No. 16;
Matches 34; Conservative 15; Mismatches 56; Indels 27; Gaps 5;

Qy 36 EEGMINISIG-----YHYPPICIGRAPGCLMPAVQNNLV-----EVTVP-NSRFT 82
Db 141 EGCHLIDSSGLVLTLYHLGNRYLTTHNCNTPWADNWLVDRGDDAEASHGLSPFGKRL 200
Qy 83 YHM-----VSGMSLRPRVNYLQ-----DFSYSRLAFRPKGTCPKEIPKGSKNTE 128
Db 201 NEMTRLGWMIDLHSVSVATKDALQISRAPVIFSHSSAYSLSLCPHRRNVPDDVLQVKNTS 260
Qy 129 VLYWEECVANSV 140
Db 261 SLVNVNFFSNFV 272

Search completed: April 9, 2002, 16:57:24
Job time: 243 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 16:57:36 ; Search time 23.68 Seconds
(without alignments)
236.897 Million cell updates/sec

Title: US-09-490-700-36

Perfect score: 840

Sequence: 1 MVTPTWMDNPIEVYVNDV.....ECVANSVVLQNEFGTIID 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	768.5	91.5	584	1 ENVI_HUMAN	P10267 homo sapien
2	89.5	10.7	529	1 RA12_MOUSE	Q9qy8 mus musculus
3	82.5	9.8	530	1 RA12_MOUSE	Q9y3p3 homo sapien
4	76	9.0	896	1 CYRB_MOUSE	P26955 mus musculus
5	75	8.9	1078	1 S24A_HUMAN	O95486 homo sapien
6	74	8.8	466	1 HN3A_HUMAN	P23512 rattus norv
7	72.5	8.6	602	1 PGH1_RAT	Q63921 rattus norv
8	72.5	8.6	1475	1 APU_THEET	P16950 t anyiopull
9	72	8.6	688	1 ENV_MMTVB	P10259 mouse mamma
10	71.5	8.5	481	1 GLC3_SOYBN	P11828 glycine max
11	71.5	8.5	1002	1 SPHR_AMEPV	P29815 amsacta moo
12	71.5	8.5	1222	1 PM5P_HUMAN	Q15155 homo sapien
13	70.5	8.4	335	1 SCN1_SCHPO	P41890 schizosacch
14	70.5	8.4	470	1 VL2_HPV54	O81023 human papill
15	70.5	8.4	598	1 P2CD_MOUSE	Q9q267 mus musculus
16	69.5	8.3	555	1 MASY_PICAN	P21360 pichia angu
17	69.5	8.3	1481	1 APU_THEET	P38939 t anyiopull
18	68.5	8.2	549	1 TREF_ECOLI	P37196 escherichia
19	68	8.1	214	1 CONA_BACSU	P14204 bacillus su
20	68	8.1	575	1 NODU_RHTR	O53515 rhicobium t
21	68	8.1	688	1 ENV_MMTVG	P03374 mouse mamma
22	68	8.1	1210	1 RP82_SCHPO	Q02061 schizosacch
23	67.5	8.0	445	1 CTB2_MOUSE	P56546 mus musculus
24	67.5	8.0	473	1 HN3A_HUMAN	P55317 homo sapien
25	67	8.0	602	1 PGH1_MOUSE	P22437 mus musculus
26	66.5	7.9	292	1 YG29_YEAST	P53242 saccharomyc
27	66.5	7.9	807	1 FSO_P_RAT	P35446 rattus norv
28	66.5	7.9	963	1 YQ36_CAEEL	Q09457 caenorhabdi
29	66.5	7.9	1091	1 MSH3_MOUSE	P13705 mus musculus
30	66.5	7.9	3396	1 PGCY_HUMAN	P13611 homo sapien
31	66	7.9	466	1 NMT_SCHPO	O43010 schizosacch
32	66	7.9	1256	1 MRP_STRSU	P32653 streptococc
33	65.5	7.8	410	1 MDPI_MOUSE	P31428 mus musculus

34	65.5	7.8	614	1 SELB_ECOLI	P14081 escherichia
35	65.5	7.8	1141	1 SYI_CAEEL	Q21926 caenorhabdi
36	65	7.7	214	1 SFS4_AQUAE	O66469 equifex aeo
37	65	7.7	338	1 FOSB_HUMAN	P33339 homo sapien
38	65	7.7	338	1 FOSB_MOUSE	P13346 mus musculus
39	65	7.7	348	1 RPC5_SCHPO	O94616 schizosacch
40	64.5	7.7	468	1 HN3A_MOUSE	P35582 mus musculus
41	64.5	7.7	599	1 PGH1_HUMAN	P23219 homo sapien
42	64.5	7.7	1409	1 COPI_DROME	P04146 drosophila
43	64.5	7.7	3358	1 PGCY_MOUSE	Q62059 mus musculus
44	64	7.6	431	1 YDPI_SCHPO	O14021 schizosacch
45	64	7.6	624	1 MTN4_MOUSE	O89029 mus musculus

ALIGNMENTS

RESULT 1	ENV1_HUMAN	STANDARD;	PRT;	584 AA.
ID	ENV1_HUMAN			
AC	P10267;			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	01-MAR-1989 (Rel. 10, Last annotation update)			
DE	RETROVIRUS-RELATED ENV POLYPROTEIN.			
GN	ENV.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87036922; PubMed=3021993;			
RA	Ono M., Yasunaga T., Miyata T., Ushikubo H.;			
RT	"Nucleotide sequence of human endogenous retrovirus genome related to			
RT	the mouse mammary tumor virus genome."			
RL	J. Virol. 60:589-598(1986).			
DR	PIR; E24483; VCHUER.			
KW	Glycoprotein; Coat protein; Polyprotein.			
FT	CARBOHYD 14 14			
FT	CARBOHYD 39 39			
FT	CARBOHYD 159 159			
FT	CARBOHYD 240 240			
FT	CARBOHYD 257 257			
FT	CARBOHYD 346 346			
FT	CARBOHYD 392 392			
FT	CARBOHYD 439 439			
FT	CARBOHYD 451 451			
FT	CARBOHYD 470 470			
SQ	SEQUENCE 584 AA; 66076 MW; FD64418F1C619166 CRC64;			

Query Match 91.5%; Score 768.5; DB 1; Length 584;
Best Local Similarity 94.7%; Pred. No. 2.3e-69;
Matches 142; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

Qy	4	PVTWMDNPIEVYVNDVSVVVGPTDRCPCAKPEEGMINISIGYHYPPICLGRAPGCLMP	63
Db	1	PVTWMDNPIEVYVNDVSVVVGPTDRCPCAKPEEGMINISIGYHYPPICLGRAPGCLMP	60
Qy	64	AVQNLWVEVTPVSPNSRFTYHVMVSGMSLRPRVNYLQDFSYQSRSLKFRPKGTCPEIKPG	123
Db	61	AVQNLWVEVTPVSPNSRFTYHVMVSGMSLRPRVNYLQDFSYQSRSLKFRPKGTCPEIKPE	120
Qy	124	SKNTEVLWEECVANSVVLQNEFGTIID	153
Db	121	SKNTEVLWEECVANSVVLQNEFGTIID	149

RESULT 2	RA12_MOUSE	STANDARD;	PRT;	529 AA.
ID	RA12_MOUSE			
AC	Q9QY8;			

FT DOMAIN 416 440 ZINC FINGER-LIKE.
SQ SEQUENCE 1078 AA; 118791 MW; 9A3214F52D28FD3F CRC64;

Query Match 8.9%; Score 75; DB 1; Length 1078;
Best Local Similarity 26.1%; Pred.No. 11;
Matches 31; Conservative 13; Mismatches 41; Indels 34; Gaps 5;

OY 44 SIGVHYPPICLRAPGCLMPAVQNMVLVEPTVSPNSRFTY-----HM--- 85
| : | | | | : | | | | : | | | : | | | : | | | :
DB 272 SVGVSYF-----SLPFGYQNTTPPGATGVPSSLNYPGSAQTPTPLGANHLTTS 322

OY 86 VSGMSLPR-----VNKLQDSFYORSKFRPKGTCKPEIKPGSKNTEVILWEECVANSV 140
| : | | | | : | | | | : | | | : | | | : | | | :
DB 323 MSGLSLQPEGLRVNLLQERNMLPSTPLKPPVNLHEDIOKLNCNPFL---RCTLTSTI 378

RESULT 6
HN3A_RAT STANDARD; PRT; 466 AA.

AC R23512;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HEPATOCYTE NUCLEAR FACTOR 3-ALPHA (HNF-3A).
GN HNF3A OR TCF3A OR TCF-3A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=101116;
RX [1]
RY SEQUENCE FROM N.A., AND SEQUENCE OF 313-337 AND 413-434.
RC TISSUE=Liver;
RX MEDLINE=91032994; PubMed=2227418;
RA Lai E., Prezioso V.R., Smith E., Litvin O., Costa R.H.,
RA Darnell J.E. Jr.;
RT "HNF-3A, a hepatocyte-enriched transcription factor of novel
RT structure is regulated transcriptionally";
RL Genes Dev. 4:1427-1436(1990).
CC -!- FUNCTION: TRANSCRIPTION ACTIVATOR FOR A NUMBER OF LIVER GENES
CC SUCH AS AFP, ALBUMIN, TYROSINE AMINOTRANSFERASE, PERCK, ETC.
CC INTERACTS WITH THE CIS-ACTING REGULATORY REGIONS OF THESE GENES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: LIVER.
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

DR EMBL; X55955; CAA39418.1; -.
DR PIR; A36674; A36674.
DR TRANSFAC; T00371; -.
DR InterPro; IPR001766; Fork head.
DR Pfam; PF00250; Fork head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation; Activator.
FT DOMAIN 251 288 ESSENTIAL FOR DNA BINDING.
FT DNA_BIND 169 260 FORK-HEAD.
FT SEQUENCE 466 AA; 48774 MW; 06555BA74BD9B7DC CRC64;

Query Match 8.8%; Score 74; DB 1; Length 466;
Best Local Similarity 25.2%; Pred.No. 5.2;
Matches 29; Conservative 13; Mismatches 57; Indels 16; Gaps 3;

[illegible]


```
RX MEDLINE=89296500; PubMed=2740231;
RA Cho T.-J., Nielsen N.C.;
RT "The glycinin Gv3 gene from soybean.";
RL Nucleic Acids Res. 17:4388-4388(1989).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=92393391; PubMed=2485233;
RA Nielsen N.C., Dickinson C.D., Cho T.-J., Thanh V.H., Scallan B.J.,
RT Fischer R.L., Sims T.L., Drews G.N., Goldberg R.B.;
RL "Characterization of the glycinin gene family in soybean.";
PL Plant Cell 1:313-328(1989).
CC -!- FUNCTION: GLYCININ IS THE MAJOR SEED STORAGE PROTEIN OF SOYBEAN.
CC -!- SUBUNIT: HEXAMER; EACH SUBUNIT IS COMPOSED OF AN ACIDIC AND A
CC BASIC CHAIN DERIVED FROM A SINGLE PRECURSOR AND LINKED BY A
CC DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE 11S SEED STORAGE PROTEIN (GLOBULINS)
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15123; CAA33217.1;
DR PIR; S04605; S04605;
DR InterPro; IPR000459; Seedstore_11s.
DR Pfam; PF00190; Seedstore_11s; 1.
DR PRINTS; PR00439; 11SGLOBULIN.
DR PROSITE; PS00305; 11S-SEED-STORAGE; 1.
KW Seed storage protein; Multigene family; Signal.
FT SIGNAL 1 21
FT CHAIN 22 296 A SUBUNIT.
FT CHAIN 297 476 B SUBUNIT.
FT PROPEP 477 481
FT DISULFID 107 303 INTERCHAIN (BY SIMILARITY).
SQ SEQUENCE 481 AA; 54241 MW; 5F3C3148DF6241A7 CRC64;

Query Match 8.5%; Score 71.5; DB 1; Length 481;
Best Local Similarity 18.1%; Pred. No. 9.6;
Matches 31; Conservative 26; Mismatches 57; Indels 57; Gaps 4;

Qy 6 TWMDNPIEVVNDVWVPGPTDRCAPKEE-----EGMMINIS 44
Db 83 SYTNAPEIYIQQSGIFGMIFPGCPSTFEFPQKQSRPQDRHQKIYHFRGDLIAVP 142
Qy 45 IGYHYPPICLGRAPCCLMPAVQNNLV---EVPTVSPNSRFTYHMYSGMSLRPRVNYLQDF 101
Db 143 TGFAY-----WMYNNEDTPVAVSLIDTNSFQNLQDQMPRRFYLAGN 184
Qy 102 SYQSLKFRPGKTCPEIKPGSKNTEVLWEECVANSVILQNNFEFTII 152
Db 185 QBEFLQYQPKQKQGGTQSKGKROQ-----EENEGGSIL 220

RESULT 11
SPHR_AMEPV
ID SPHR_AMEPV STANDARD; PRT; 1002 AA.
AC P29815;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE SPHEROIDIN.
GN G5.
OS Amsacta moorei entomopoxvirus (AmePV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=28321;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
```

```
RX MEDLINE=92185464; PubMed=1545219;
RA Banville M., Dumas F., Trifiro S., Arif B., Richardson C.;
RT "The predicted amino acid sequence of the spheroidin protein from
RT Amsacta moorei entomopoxvirus; lack of homology between major
RT occlusion body proteins of different poxviruses.";
RL J. Gen. Virol. 73:559-566(1992).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92046310; PubMed=1942245;
RA Hall R.L., Moyer R.W.;
RT "Identification, cloning, and sequencing of a fragment of Amsacta
RT moorei entomopoxvirus DNA containing the spheroidin gene and three
RT vaccinia virus-related open reading frames.";
RL J. Virol. 65:6516-6527(1991).
CC -!- FUNCTION: MAJOR COMPONENT OF VIRAL OCCLUSION BODIES, THE
CC PROTECTIVE COMPLEXES IN WHICH THE VIRIONS ARE EMBEDDED IN THE
CC CYTOPLASM OF THEIR INSECT HOSTS.
CC -!- SUBUNIT: MAY FORM DISULFIDE BONDS LINKED AGGREGATES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M75889; AAA42378.1;
DR EMBL; M77182; AAA42383.1;
DR PIR; J01436; PIVZAM.
KW Viral occlusion body; Acetylation; Late protein.
FT INIT_MET 0
FT MOD_RES 1 1 ACETYLATION.
SQ SEQUENCE 1002 AA; 114738 MW; DC0D61B157645F85 CRC64;

Query Match 8.5%; Score 71.5; DB 1; Length 1002;
Best Local Similarity 25.0%; Pred. No. 22;
Matches 26; Conservative 12; Mismatches 37; Indels 29; Gaps 5;

Qy 60 CLMPAVQ-----NWLVVEPTVSPNSRFTYHMYSG---MSLRPRVNYLQDFSYQSLKFR 110
Db 440 CLKPKVPKNLRMGWILDCDT---SRFIKHAQSGDDLDLDVRLN----- 481
Qy 111 PKGKTCPEK-IPKGSKNTEVLWEECVANSVILQNNFEFTII 153
Db 482 -RNDICLQAQKQHTVNVILLEAYNTYPNCTLSLGNRRNNVFD 524

RESULT 12
PM5P_HUMAN
ID PM5P_HUMAN STANDARD; PRT; 1222 AA.
AC Q15155; P78421;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN PM5 PRECURSOR.
GN PM5
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 33-1222 FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=92128984; PubMed=1310294;
RA Templeton N.S., Rodgers L.A., Levy A.T., Ting K.-L., Krutzsch H.C.,
RA Liotta L.A., Stetler-Stevenson W.G.;
RT "Cloning and characterization of a novel human cDNA that has DNA
RT similarity to the conserved region of the collagenase gene family.";
RL Genomics 12:175-176(1992).
RN [2]
RP SEQUENCE OF 1-1148 FROM N.A.
```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U37488; AAA79192.1; -;
 DR InterPro; IPR000784; Late_L2;
 DR Pfam; PF00513; late_protein_L2; 1.
 KW Coat protein; Late protein.
 SQ SEQUENCE 470 AA; 50492 MW; CA0BFB5D8D05E5DF CRC64;

Query Match 8.4%; Score 70.5; DB 1; Length 470;
 Best Local Similarity 25.7%; Pred. No. 12;
 Matches 26; Conservative 16; Mismatches 42; Indels 17; Gaps 4;
 QY 13 EYVNDVSVMVPGPTDDRCAPKPEEGMINISIGVHYPPICIGRAPGCLMPAVON-WLVE 71
 DB 233 QPVQDPAFLQOPSSLIYDNPVYEG---NPDVTLHFQPTIHNAPD---PAFMDFIALH 286
 QY 72 VPTVSP-----NSRFTYHVMGSLRPRVNYLQDFS 102
 DB 287 RPALTRRGVVRYSRVGDRATLHTRSGIQLKPRVHFFQDLS 327

RESULT 15
 P2CD_MOUSE
 ID P2CD_MOUSE STANDARD; PRT; 598 AA.
 AC Q90Z67;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN PHOSPHATASE 2C DELTA ISOFORM (EC 3.1.3.16) (PP2C-DELTA) (P53-
 DE INDUCED PROTEIN PHOSPHATASE 1) (PROTEIN PHOSPHATASE MAGNESIUM-
 DE DEPENDENT 1 DELTA).
 GN PPM1D OR WIP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SVJ;
 RA Choi J., Appella E., Donehower L.;
 RT "The structure and expression of the murine wild-type p53-induced
 RT phosphatase 1 (wip1) gene."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RC -1- FUNCTION: MIGHT CONTRIBUTE TO GROWTH INHIBITORY PATHWAYS ACTIVATED
 CC IN RESPONSE TO DNA DAMAGE IN A P53-DEPENDENT MANNER.
 CC -1- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN +
 CC ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).
 CC -1- COFACTOR: BINDS TWO MAGNESIUM OR MANGANESE IONS (BY SIMILARITY).
 CC -1- INDUCTION: BY P53.
 CC -1- SIMILARITY: BELONGS TO THE PP2C FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AF200464; AAF09251.1; -;
 DR MGD; MGI:1858214; Ppm1d.
 DR InterPro; IPR000222; PP2C.
 DR InterPro; IPR003589; PP2C_catalytic.
 DR InterPro; IPR001932; PP2C_domain.
 DR InterPro; IPR003588; PP2C_sig.
 DR Pfam; PF00481; PP2C; 1.
 DR SMART; SM00332; PP2Cc; 1.
 DR SMART; SM00331; PP2C_SIG; 1.
 DR PROSITE; PS01032; PP2C; 1.
 KW Hydrolase; Magnesium; Manganese; Multigene family.
 FT DOMAIN 15 371 PP2C-LIKE.

SQ SEQUENCE 598 AA; 65722 MW; 4DB70B5D48539435 CRC64;
 Query Match 8.4%; Score 70.5; DB 1; Length 598;
 Best Local Similarity 22.9%; Pred. No. 16;
 Matches 50; Conservative 23; Mismatches 74; Indels 71; Gaps 10;
 QY 3 TPVTWMDNPPIEVYV---NDSVM--VPGPTDDRCAPKPEEGMINISIGVHYPPICIGRA 57
 DB 289 TSVHTLDPKHKYIILSGDGLNNVP-PQDAISMCDQDEBKYLNGEQGQSCAKMLVNR 347
 QY 58 PG-----CLMPAVQNW-----LVEVPTVSPNSRFTYHVMVSG 88
 DB 348 LGRWRQRLRADNTSAIVICISPEVDNQGNTNEDELFLNLTDSPTY--NSQETCVMTSS 405
 QY 89 MSLRPRVNYLQDFSQ-----RSLKFRPKGKTCPEIKPG-----SKNTE 128
 DB 406 PSSTPPIKSPPEEDAMPRLSSKDHIPALVRSNAFSEKFLVPAEIAIARGNIQTIVVMTSKDSE 465
 QY 129 VLVVEECV-----ANSVILQNNFEFTIID 153
 DB 466 TLE-ENCPKALTIRIHDSLNNTLSVGLIPTNSTNTIMD 502

Search completed: April 9, 2002, 17:07:01
 Job time: 565 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 16:48:46 ; Search time 72.35 Seconds
(without alignments)
156.644 Million cell updates/sec

Title: US-09-490-700-36

Perfect score: 840

Sequence: 1 MVTPTWMDNPIEVVDSV.....ECVANSVILQNEFGTIID 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	840	100.0	153	20 AAW95692	Human endogenous r
2	840	100.0	153	20 AAW97745	Human endogenous r
3	840	100.0	181	20 AAW95694	Human endogenous r
4	840	100.0	181	20 AAW97747	Human endogenous r
5	840	100.0	561	20 AAW95693	Human endogenous r
6	840	100.0	561	20 AAW97746	Human endogenous r
7	220	26.2	48	22 AAM14052	Peptide #486 encod
8	220	26.2	48	22 AAM26458	Peptide #495 encod
9	220	26.2	48	22 AAM01792	Peptide #474 encod
10	72.5	8.6	1475	11 AAR08221	Recombinant alpha
11	72	8.6	688	20 AAY41142	Mouse mammary tumo

12	71.5	8.5	1003	13 AAR29648	AmEPV Spheroidin p
13	71.5	8.5	1003	15 AAR55576	AmEPV spheroidin.
14	71.5	8.5	1003	19 AAW41301	AmEPV entomopoxvir
15	71.5	8.5	1003	20 AAY30169	Spheroidin protein
16	71	8.5	725	22 AAY72914	E. coli-ironec ext
17	70.5	8.4	410	20 AAY49062	Mouse membrane dip
18	70.5	8.4	1248	22 AAM39055	Human polyprotein
19	70	8.3	252	21 AAG47392	Human polyprotein
20	70	8.3	256	21 AAG31623	Arabidopsis thalia
21	70	8.3	260	21 AAG44355	Arabidopsis thalia
22	69.5	8.3	1464	17 AAR88469	Feline infectious
23	69	8.2	185	21 AAG31812	Arabidopsis thalia
24	69	8.2	196	21 AAG31811	Arabidopsis thalia
25	69	8.2	204	21 AAG31810	Arabidopsis thalia
26	69	8.2	468	19 AAW74797	Human secreted pro
27	69	8.2	549	20 AAY60008	Human endometrium
28	69	8.2	1502	22 AAM39273	Human polyprotein
29	69	8.2	1565	22 AAM41059	Human polyprotein
30	68	8.1	142	20 AAY33542	B. juncea CBF homo
31	67.5	8.0	95	22 AAG74016	Human colon cancer
32	67.5	8.0	484	18 AAW19006	Feline herpes viru
33	66.5	7.9	505	20 AAY36807	Protein involved i
34	66.5	7.9	741	21 AAB43485	Human cancer assoc
35	66.5	7.9	807	14 AAR44241	F-spondin (FP5-9).
36	66.5	7.9	807	20 AAY04264	Rat vascular smoot
37	66.5	7.9	807	21 AAB13074	Vascular endotheli
38	66.5	7.9	2409	12 AAR12609	Versican. Homo sa
39	66	7.9	488	22 AAU14847	Novel bone marrow
40	66	7.9	704	21 AAY75590	Neisseria meningit
41	66	7.9	704	21 AAY75591	Neisseria meningit
42	66	7.9	930	21 AAB18212	Plasmodium falcipa
43	66	7.9	1256	13 AAR27746	Muramidase release
44	65.5	7.8	410	15 AAP53920	Dehydropeptidase-1
45	65.5	7.8	3023	17 AAR94462	Hepatitis C virus

ALIGNMENTS

RESULT 1
AAW95692
ID AAW95692 standard; Protein: 153 AA.
AC AAW95692;
XX
XX
DT 08-JUN-1999 (first entry)
XX
DE Human endogenous retrovirus IDMK1.2-22 env protein.
XX
IDDM; insulin-dependent diabetes mellitus; endogenous retrovirus;
KW SAG; superantigen; provirus; autoimmune disease; type 1 diabetes;
KW diagnosis; env; envelope.
XX
OS Homo sapiens.
XX
XX
PN EP893691-A1.
XX
PD 27-JAN-1999.
XX
PF 23-JUL-1997; 97EP-0401773.
XX
PR 23-JUL-1997; 97EP-0401773.
XX
PA (MACH/) MACH B F.
XX
PI Conrad B, Mach B;
XX
DR WPI; 1999-097928/09.
DR N-PSDB; AAX07514.
XX
PT Diagnosing human autoimmune disease by detecting retrovirus with
PT superantigen activity - new retrovirus associated with type 1
PT diabetes, its proviral DNA, and related vectors, transformed cells,

PT proteins, antibodies and specific binding agents, used for treating
 XX or preventing autoimmune disease
 PS Claim 31; Fig 7D; 92pp; English.
 XX
 CC The sequence is that of an insulin-dependent diabetes mellitus
 CC associated human endogenous retrovirus (IDDMK1.2-22) env protein.
 CC The retrovirus has Superantigen (SAG) activity. It can be used
 CC as part of a method is specifically used to diagnose type 1 diabetes
 CC mellitus. Modified proteins expressed by the retroviral sequence
 CC (without SAG activity but still able to induce an immune response)
 CC are useful in vaccines to treat or prevent SAG-related autoimmune
 CC disease; nucleic acid sequences encoding (modified) SAG can be used
 CC similarly to treat such diseases. Retroviral-encoded SAG are important
 CC in pathogenesis of autoimmune disease, probably by activating
 CC autoreactive T cells. The method is very specific (it can differentiate
 CC between expressed and non-expressed viral nucleic acids) and can be used
 CC even where the pathogen is an ubiquitous endogenous retrovirus. Blood
 CC or plasma samples can be tested without extensive preparation and
 CC diagnosis can be made before clinical signs are apparent, allowing
 CC early intervention before severe tissue damage has occurred.
 XX
 SQ Sequence 153 AA;

Query Match 100.0%; Score 840; DB 20; Length 153;
 Best Local Similarity 100.0%; Pred. No. 8.5e-88;
 Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MVTPTWMDNPIEVYVNDVSVVPGPTDDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 60
 Db 1 mvtptwmdnplievynvndsvvpgptddrcapkeegmminisigyhyppicigrapgc 60
 Qy 61 LMPAVQNWLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPEI 120
 Db 61 lmpavqnlwlvvptvpsnsrftyhmvsgrmslrprvnylqdfsyqrsikfrpkgtcpkei 120
 Qy 121 PKGSKNTEVLWEECVANSVVIQNNFGTTID 153
 Db 121 pkgskntevlweecvansvvlqnnfgttiid 153

RESULT 2
 AAW97745
 ID AAW97745 standard; Protein; 153 AA.
 XX
 AC AAW97745;
 XX
 DT 21-MAY-1999 (first entry)
 XX
 DE Human endogenous retrovirus IDDKK1.2-22 envelope protein.
 XX
 KW HERV; IDDKK1.2-22; superantigen; SAG; antigen; IDDM;
 KW insulin-dependent diabetes mellitus; autoimmune disease; diagnosis;
 KW therapy; vaccine; envelope protein; env gene.
 XX
 OS Human endogenous retrovirus.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 17...19
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 42..45
 FT /note= "Asn is N-glycosylated"
 XX
 PN W09905527-A2.
 XX
 XX 04-FEB-1999.
 XX
 XX 22-JUL-1998; 98WO-EP04936.
 XX
 XX 23-JUL-1997; 97EP-0401773.
 PR 22-JUL-1997; 97EP-0112482.
 XX

PA (MEDI-) MEDIGEN SA.
 XX Conrad B, Mach B;
 XX WPI; 1999-143118/12.
 DR N-PSDB; AAX07189, AAX07189.
 XX
 PT New isolated human endogenous retrovirus - used to develop products
 PT for the diagnosis, prevention and treatment of autoimmune disease,
 PT particularly insulin dependent diabetes mellitus
 XX
 PS Claim 6; Fig 7D; 165pp; English.
 XX
 CC This is the envelope protein (Env) of a new human endogenous
 CC retrovirus (HERV), designated IDDMK1.2-22, that has been identified
 CC as the source of superantigen (SAG) activity in insulin-dependent
 CC diabetes mellitus (IDDM) patients. The endogenous retrovirus is
 CC ubiquitous in the human genome but is only expressed in diabetic
 CC individuals. The HERV encodes SAG activity within the env gene.
 CC A claimed process for the diagnosis, including the pre-symptomatic
 CC diagnosis, of a human autoimmune disease associated with a HERV
 CC having SAG activity comprises specifically detecting in a
 CC biological sample either: (a) the mRNA of an expressed HERV having
 CC SAG activity (especially IDDMK1.2-22 5'LTR, 3' long terminal
 CC repeat, env or pol); (b) a protein or peptide expressed by the HERV
 CC (see AAW97745-48); (c) antibodies specific to the proteins expressed
 CC by the HERV; or (d) SAG activity specifically associated with the
 CC HERV. Products of the invention can be used to identify substances
 CC capable of blocking transcription or translation of SAG-encoding
 CC nucleic acid sequences, useful in therapy and/or prevention of
 CC autoimmune disease associated with the SAG. A nucleic acid encoding
 CC human retroviral SAG can be used as a DNA vaccine. Expression of
 CC the endogenous SAG in IDDM suggests a general model according to
 CC which self SAG-driven and systemic activation of autoreactive T
 CC cells leads to organ-specific autoimmune disease.
 XX
 SQ Sequence 153 AA;

Query Match 100.0%; Score 840; DB 20; Length 153;
 Best Local Similarity 100.0%; Pred. No. 8.5e-88;
 Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MVTPTWMDNPIEVYVNDVSVVPGPTDDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 60
 Db 1 mvtptwmdnplievynvndsvvpgptddrcapkeegmminisigyhyppicigrapgc 60
 Qy 61 LMPAVQNWLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPEI 120
 Db 61 lmpavqnlwlvvptvpsnsrftyhmvsgrmslrprvnylqdfsyqrsikfrpkgtcpkei 120
 Qy 121 PKGSKNTEVLWEECVANSVVIQNNFGTTID 153
 Db 121 pkgskntevlweecvansvvlqnnfgttiid 153

RESULT 3
 AAW95694
 ID AAW95694 standard; Protein; 181 AA.
 XX
 AC AAW95694;
 XX
 DT 08-JUN-1999 (first entry)
 XX
 DE Human endogenous retrovirus IDDMK1.2-22 env/fs (sag) protein.
 XX
 KW IDDM; insulin-dependent diabetes mellitus; endogenous retrovirus;
 KW SAG; superantigen; provirus; autoimmune disease; type 1 diabetes;
 KW diagnosis; env; envelope.
 XX
 OS Homo sapiens.
 XX
 PN EP893691-A1.

```

XX PD 27-JAN-1999.
XX PH
XX FT 23-JUL-1997; 97EP-04011773.
XX PR 23-JUL-1997; 97EP-04011773.
XX PA (MACH/) MACH B F.
XX PI Conrad B, Mach B;
XX PD WPI; 1999-097928/09.
XX DR N-PSDB; AAX07516.
XX
XX Diagnosing human autoimmune disease by detecting retrovirus with
XX PT superantigen activity - new retrovirus associated with type 1
XX PT diabetes, its proviral DNA, and related vectors, transformed cells,
XX PT proteins, antibodies and specific binding agents, used for treating
XX PT or preventing autoimmune disease
XX
XX Claim 31; Fig 7G; 92pp; English.
XX
XX The sequence is that of an insulin-dependent diabetes mellitus
XX CC associated human endogenous retrovirus (IDMK1.2-22) env/fs protein.
XX CC The retrovirus has Superantigen (Sag) activity. It can be used
XX CC as part of a method is specifically used to diagnose type 1 diabetes
XX CC mellitus. Modified proteins expressed by the retroviral sequence
XX CC (without Sag activity but still able to induce an immune response)
XX CC are useful in vaccines to treat or prevent Sag-related autoimmune
XX CC disease; nucleic acid sequences encoding (modified) Sag can be used
XX CC similarly to treat such diseases. Retroviral-encoded Sag are important
XX CC in pathogenesis of autoimmune disease, probably by activating
XX CC autoreactive T cells. The method is very specific (it can differentiate
XX CC between expressed and non-expressed viral nucleic acids) and can be used
XX CC even where the pathogen is an ubiquitous endogenous retrovirus. Blood
XX CC or plasma samples can be tested without extensive preparation and
XX CC diagnosis can be made before clinical signs are apparent, allowing
XX CC early intervention before severe tissue damage has occurred.
XX
XX Sequence 181 AA;

Query Match 100.0%; Score 840; DB 20; Length 181;
Best Local Similarity 100.0%; Pred. No. 1.1e-87;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVTPTVMDNPIEVYVNDVSVVPGPTDDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 60
DB 1 mvtptvmdnpienvydvndsvvpgptddrcapkeegmmminisigyhyppiclgrapgc 60

QY 61 LMPAVQNLVVEPTVSPNSRFTYHMGSLRPRVNYLQDFSYQSLKFRPKGKTCPKKEI 120
DB 1 mpavqnlvveptvspnsrftyhmvgmslrprvnylqdfsyqslkfrpkgktpckei 120

QY 121 PKGSKNTEVLVWEECVANSVVILQNNFEFGTIID 153
DB 1 pkgsknteavlweecvansvvilqnnfeftiid 153

RESULT 4
AAW97747
ID AAW97747 standard; Protein; 181 AA.
XX
XX AAW97747;
XX
XX 21-MAY-1999 (first entry)
XX
XX Human endogenous retrovirus IDDKK1.2-22 Env/F-S (Sag).
XX
XX HPRV; IDDKK1.2-22; superantigen; Sag; antigen; IDDM;
XX CC insulin-dependent diabetes mellitus; autoimmune disease; diagnosis;
XX KW therapy; vaccine; envelope protein; env gene.
XX

```

```

OS Human endogenous retrovirus.
XX
XX Key Location/Qualifiers
XX FT Modified-site 17..19
XX FT Modified-site /note= "Asn is N-glycosylated"
XX FT Modified-site 42..45
XX FT Modified-site /note= "Asn is N-glycosylated"
XX
XX WO9905527-A2.
XX
XX 04-FEB-1999.
XX
XX 22-JUL-1998; 98WO-EP04926.
XX
XX 23-JUL-1997; 97EP-04011773.
XX PR 22-JUL-1997; 97EP-0112482.
XX
XX (MEDI-) MEDIGEN SA.
XX
XX Conrad B, Mach B;
XX
XX WPI; 1999-143118/12.
XX DR N-PSDB; AAX07191.
XX
XX New isolated human endogenous retrovirus - used to develop products
XX PT for the diagnosis, prevention and treatment of autoimmune disease,
XX PT particularly insulin dependent diabetes mellitus
XX
XX Claim 6; Fig 7G; 165pp; English.
XX
XX This is the envelope (Env)/FS polypeptide of a new human endogenous
XX CC retrovirus (HERV), designated IDDMK1.2-22, that has been identified
XX CC as the source of superantigen (Sag) activity in insulin-dependent
XX CC diabetes mellitus (IDDM) patients. The endogenous retrovirus is
XX CC ubiquitous in the human genome but is only expressed in diabetic
XX CC individuals. The HERV encodes Sag activity within the env gene.
XX CC A claimed process for the diagnosis, including the pre-symptomatic
XX CC diagnosis, of a human autoimmune disease associated with a HERV
XX CC having Sag activity comprises specifically detecting in a
XX CC biological sample either: (a) the mRNA of an expressed HERV having
XX CC Sag activity (especially IDDMK1.2-22 5'LTR, 3' long terminal
XX CC repeat, env or pol); (b) a protein or peptide expressed by the HERV
XX CC (see AAW97745-48); (c) antibodies specific to the proteins expressed
XX CC by the HERV; or (d) Sag activity specifically associated with the
XX CC HERV. Products of the invention can be used to identify substances
XX CC capable of blocking transcription or translation of Sag-encoding
XX CC nucleic acid sequences, useful in therapy and/or prevention of
XX CC autoimmune disease associated with the Sag. A nucleic acid encoding
XX CC human retroviral Sag can be used as a DNA vaccine. Expression of
XX CC the endogenous Sag in IDDM suggests a general model according to
XX CC which self Sag-driven and systemic activation of autoreactive T
XX CC cells leads to organ-specific autoimmune disease.
XX
XX Sequence 181 AA;

```

```

Query Match 100.0%; Score 840; DB 20; Length 181;
Best Local Similarity 100.0%; Pred. No. 1.1e-87;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVTPTVMDNPIEVYVNDVSVVPGPTDDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 60
DB 1 mvtptvmdnpienvydvndsvvpgptddrcapkeegmmminisigyhyppiclgrapgc 60

QY 61 LMPAVQNLVVEPTVSPNSRFTYHMGSLRPRVNYLQDFSYQSLKFRPKGKTCPKKEI 120
DB 1 mpavqnlvveptvspnsrftyhmvgmslrprvnylqdfsyqslkfrpkgktpckei 120

QY 121 PKGSKNTEVLVWEECVANSVVILQNNFEFGTIID 153
DB 1 pkgsknteavlweecvansvvilqnnfeftiid 153

```

```
RESULT 5
AAW95693
ID AAW95693 standard; Protein: 561 AA.
XX
AC AAW95693;
XX
DT 08-JUN-1999 (first entry)
XX
DE Human endogenous retrovirus IDDMK1.2-22 env/fs-sag protein.
XX
KW IDDM; insulin-dependent diabetes mellitus; endogenous retrovirus;
KW SAG; superantigen; provirus; autoimmune disease; type 1 diabetes;
KW diagnosis; env; envelope.
XX
OS Homo sapiens.
XX
PN EP893691-A1.
XX
PD 27-JAN-1999.
XX
PF 23-JUL-1997; 97EP-0401773.
XX
PR 23-JUL-1997; 97EP-0401773.
XX
PA (MACH/) MACH B F.
XX
PI Conrad B, Mach B;
XX
DR WPI: 1999-097928/09.
DR N-PSDB; AAX07515.
XX
Diagnosing human autoimmune disease by detecting retrovirus with
superantigen activity - new retrovirus associated with type 1
diabetes, its proviral DNA, and related vectors, transformed cells,
proteins, antibodies and specific binding agents, used for treating
or preventing autoimmune disease
XX
PS Claim 31; Fig 7E; 92pp; English.
XX
The sequence is that of an insulin-dependent diabetes mellitus
associated human endogenous retrovirus (IDDMK1.2-22) env/fs-sag protein.
XX
The retrovirus has Superantigen (SAG) activity. It can be used
as part of a method is specifically used to diagnose type 1 diabetes
mellitus. Modified proteins expressed by the retroviral sequence
(without SAG activity but still able to induce an immune response)
are useful in vaccines to treat or prevent SAG-related autoimmune
disease; nucleic acid sequences encoding (modified) SAG can be used
similarly to treat such diseases. Retroviral-encoded SAG are important
in pathogenesis of autoimmune disease, probably by activating
autoreactive T cells. The method is very specific (it can differentiate
between expressed and non-expressed viral nucleic acids) and can be used
even where the pathogen is an ubiquitous endogenous retrovirus. Blood
or plasma samples can be tested without extensive preparation and
diagnosis can be made before clinical signs are apparent, allowing
early intervention before severe tissue damage has occurred.
XX
SQ Sequence 561 AA;

Query Match 100.0%; Score 840; DB 20; Length 561;
Best Local Similarity 100.0%; Pred. No. 4.6e-87;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVTPTWMDNPIEVYVNDVWVPGTDDRCAPKPEEGMINISIGYHYPPICLGRAPGC 60
|||||
Db 1 mvtptwmdnpieryyndsvwpptddrcapakpeegminisigyhyppicigrapgc 60
|||||

Qy 61 LMPAVQNVLVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQRSLKFRPKGKTCPKEI 120
|||||
Db 61 lmpavqnlvvevptvpsnrftymvsgmslrprvnylqdfsyqrsllkfrpkgtcpkei 120
|||||

Qy 121 PKGSKNTEVLVWEECVANSVVLQNNEFGTIIID 153
|||||
```

```
Db 121 pkgskntevlvweecvansvvlqnnfegtliid 153
RESULT 6
AAW97746
ID AAW97746 standard; Protein: 561 AA.
XX
AC AAW97746;
XX
DT 21-MAY-1999 (first entry)
XX
DE Human endogenous retrovirus IDDMK1.2-22 envelope protein.
XX
KW HERV; IDDMK1.2-22; superantigen; SAG; antigen; IDDM;
KW insulin-dependent diabetes mellitus; autoimmune disease; diagnosis;
KW therapy; vaccine; envelope protein; env gene.
XX
OS Human endogenous retrovirus.
XX
Key Location/Qualifiers
FH Modified-site 17..19
FT /note= "Asn is N-glycosylated"
FT Modified-site 42..45
FT /note= "Asn is N-glycosylated"
XX
WO9905527-A2.
XX
PD 04-FEB-1999.
XX
PF 22-JUL-1998; 98WO-EP04926.
XX
PR 23-JUL-1997; 97EP-0401773.
XX
PR 22-JUL-1997; 97EP-0112482.
XX
PA (MEDI-) MEDIGEN SA.
XX
PI Conrad B, Mach B;
XX
DR WPI: 1999-143118/12.
XX
New isolated human endogenous retrovirus - used to develop products
for the diagnosis, prevention and treatment of autoimmune disease,
particularly insulin dependent diabetes mellitus
XX
PS Claim 6; Fig 7E; 165pp; English.
XX
This is the envelope protein (Env) of a new human endogenous
retrovirus (HERV), designated IDDMK1.2-22, that has been identified
as the source of superantigen (SAG) activity in insulin-dependent
diabetes mellitus (IDDM) patients. The endogenous retrovirus is
ubiquitous in the human genome but is only expressed in diabetic
individuals. The HERV encodes SAG activity within the env gene.
XX
A claimed process for the diagnosis, including the pre-symptomatic
diagnosis, of a human autoimmune disease associated with a HERV
having SAG activity comprises specifically detecting in a
biological sample either: (a) the mRNA of an expressed HERV
SAG activity (especially IDDMK1.2-22 5'LTR, 3' long terminal
repeat, env or pol); (b) a protein or peptide expressed by the HERV
(see AAW97745-48); (c) antibodies specific to the proteins expressed
by the HERV; or (d) SAG activity specifically associated with the
HERV. Products of the invention can be used to identify substances
capable of blocking transcription or translation of SAG-encoding
nucleic acid sequences, useful in therapy and/or prevention of
autoimmune disease associated with the SAG. A nucleic acid encoding
human retroviral SAG can be used as a DNA vaccine. Expression of
the endogenous SAG in IDDM suggests a general model according to
which self SAG-driven and systemic activation of autoreactive T
cells leads to organ-specific autoimmune disease.
XX
SQ Sequence 561 AA;

Query Match 100.0%; Score 840; DB 20; Length 561;
```

Best Local Similarity 100.0%; Pred. No. 4.6e-87;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVTPTVMDNPIEVYVNDSDVWVWPCPTDRCPCAKPEEEGMMINISIGYHYPPICLGRAPGC 60
Db 1 mvtptvmdnpienvyvnsvwvwpptddrcpcakpeeeegmmminisigyhyppicigrapgc 60

Qy 61 LMPAVQNLVEVPTVSPNSRFTYHVMGMSLRPRVNTLQDFSYQSLKFRPKGKTCPKKEI 120
Db 61 lmpavqnlvevptvspnsrftyhmvsgmslrprvntlqdfsyrslkfrpkgtcpkei 120

Qy 121 PKGSKNTEVLWEECVANSVYILQNNFGTIID 153
Db 121 pkgskntevlwveevansvylqnnfgtiid 153

RESULT 7
AAM14052
ID AAM14052 standard; Protein: 48 AA.
XX
AC AAM14052;
XX
DT 12-OCT-2001 (first entry)
XX
DE Peptide #486 encoded by probe for measuring cervical gene expression.
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI: 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 27; SEQ ID No 18878; 487pp; English.
XX

The present invention relates to human single exon nucleic acid probes (SENPs; see AAI10068-AAI28459). The present sequence is a peptide encoded by one such probe. The SENPs are derived from human HeLa cells. The SENPs can be used to produce a single exon microarray, which can be used for measuring human gene expression in a sample derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

Sequence 48 AA;
Query Match 26.2%; Score 220; DB 22; Length 48;
Best Local Similarity 83.3%; Pred. No. 7.8e-18;

Matches 40; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 47 YHYPPICLGRAPGCLMPAVQNLVEVPTVSPNSRFTYHVMGMSLRPR 94
Db 1 ycyppicigrapgclmptqnlwlvvevptvsatsgftyhrvsgmslrpq 48

RESULT 8
AAM26458
ID AAM26458 standard; Protein: 48 AA.
XX
AC AAM26458;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #495 encoded by probe for measuring placental gene expression.
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI: 2001-488907/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 27; SEQ ID No 26727; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs; see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
SQ Sequence 48 AA;

Query Match 26.2%; Score 220; DB 22; Length 48;
Best Local Similarity 83.3%; Pred. No. 7.8e-18;
Matches 40; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 47 YHYPPICLGRAPGCLMPAVQNLVEVPTVSPNSRFTYHVMGMSLRPR 94
Db 1 ycyppicigrapgclmptqnlwlvvevptvsatsgftyhrvsgmslrpq 48

RESULT 9
AAM01792
ID AAM01792 standard; Protein: 48 AA.
XX
AC AAM01792;
XX
DT 09-OCT-2001 (first entry)

PT and mammalian host cells

PS Disclosure; Page 68-71; 118pp; English.

XX The sequence of the Amsacta moorei entomopoxvirus spheroidin gene

CC and its flanking regions was determined. The spheroidin gene

CC can be used as the location for the insertion of heterologous DNA

CC in insect and mammalian expression systems.

XX

SQ Sequence 1003 AA;

Query Match 8.5%; Score 71.5; DB 15; Length 1003;

Best Local Similarity 25.0%; Pred. No. 34;

Matches 26; Conservative 12; Mismatches 37; Indels 29; Gaps 5;

Qy 60 CLMPAVQ-----NWLVEVPVTPNSRFTYHMVSG---MSLRPRVNYLQDFSQKSLKFR 110

Db 441 clkpvpknrlrgwldcdt----srfikmadgsddldldvrln----- 482

Qy 111 PKGKTCPE-IPKGSKNTEVLWEECVANSVVIQNNFEFTIID 153

Db 483 -ndiclkqalkqnytnvileyantypnctlsignnrfnnvfd 525

RESULT 14

AAW41301

ID AAW41301 standard; Protein; 1003 AA.

XX

AC AAW41301;

DT 20-MAY-1998 (first entry)

XX

DE Amsacta moorei entomopoxvirus spheroidin protein G5R.

XX

KW Entomopoxvirus; spheroidin gene; AmsEPV; thymidine kinase; promoter;

KW insect control; viral vaccine.

XX

OS Amsacta moorei.

XX

PN US5721352-A.

XX

PD 24-FEB-1998.

XX

PF 22-NOV-1993; 93US-0107755.

XX

PR 22-NOV-1993; 93US-0107755.

PR 19-FEB-1991; 91US-0657584.

PR 30-JAN-1992; 92US-0827685.

PR 12-FEB-1992; 92WO-US00855.

XX

PA (UYFL) UNIV FLORIDA RES FOUND.

XX

PI Gruidl ME, Hall RL, Moyer RW;

XX

DR WPI; 1998-168476/15.

DR N-PSDB; AAV14507, AAV14517.

XX

PT New Entomopoxvirus nucleic acid sequences - used in DNA constructs

PT and vectors for expression of heterologous genes in, e.g. insect

PT cells

PS Disclosure; Columns 45-50; 55pp; English.

XX

CC This sequence is encoded by the Amsacta moorei entomopoxvirus (AmsEPV)

CC spheroidin gene, which is an example of the gene of the

CC invention, which encodes a 115 kDa protein. EPV spheroidin and

CC thymidine kinase promoters can be used in DNA constructs and vectors for

CC expression of heterologous genes in insects or mammalian cells.

CC e.g. vectors containing Bacillus thuringiensis toxin genes for use in

CC insect control, or recombinant vaccinia or swinepox viruses for use as

CC viral vaccines.

XX

SQ Sequence 1003 AA;

Query Match 8.5%; Score 71.5; DB 19; Length 1003;

Best Local Similarity 25.0%; Pred. No. 34;

Matches 26; Conservative 12; Mismatches 37; Indels 29; Gaps 5;

Qy 60 CLMPAVQ-----NWLVEVPVTPNSRFTYHMVSG---MSLRPRVNYLQDFSQKSLKFR 110

Db 441 clkpvpknrlrgwldcdt----srfikmadgsddldldvrln----- 482

Qy 111 PKGKTCPE-IPKGSKNTEVLWEECVANSVVIQNNFEFTIID 153

Db 483 -ndiclkqalkqnytnvileyantypnctlsignnrfnnvfd 525

RESULT 15

AAAY30169

ID AAAY30169 standard; Protein; 1003 AA.

XX

AC AAAY30169;

DT 28-OCT-1999 (first entry)

XX

DE Spheroidin protein encoded by Amsacta moorei entomopoxvirus genome.

XX

KW Spheroidin; Entomopoxvirus; expression system; replication;

KW heterologous gene expression; thymidine kinase; poxvirus; vaccinia;

KW swinepox virus; insect pest control; immunity.

XX

OS Amsacta moorei entomopoxvirus.

XX

PN US5935777-A.

XX

PD 10-AUG-1999.

XX

PF 17-OCT-1995; 95US-0544332.

XX

PR 17-OCT-1995; 95US-0544332.

PR 19-FEB-1991; 91US-0657584.

PR 30-JAN-1992; 92US-0827685.

PR 12-FEB-1992; 92WO-US00855.

PR 07-DEC-1992; 92US-0991867.

XX

PA (UYFL) UNIV FLORIDA RES FOUND INC.

XX

PI Gruidl ME, Hall RL, Li Y, Moyer RW;

XX

DR WPI; 1999-457596/38.

DR N-PSDB; AAZ10081.

XX

PT Novel expression system for the expression of heterologous sequences

PT in insect and mammalian host cells

XX

PS Disclosure; Column 61-68; 72pp; English.

XX

CC AAY30165-70 represent proteins encoded by open reading frames (ORFs)

CC of the Amsacta moorei entomopoxvirus spheroidin gene and flanking

CC sequences. The DNA is used to make expression systems of the invention.

CC The specification describes an Entomopoxvirus (EPV) expression system

CC that is capable of directing the replication and expression of a

CC heterologous gene in a selected host cell. The expression system

CC comprises an EPV promoter sequence operably linked to the selected

CC heterologous gene sequence. The expression system is used for the

CC expression of heterologous sequences and the production of selected

CC proteins in insect and mammalian host cells e.g. human, rodent and

CC primate cells. EPV thymidine kinase and spheroidin genes can also be

CC used in vertebrate poxviruses such as vaccinia and swinepox virus. The

CC expression vectors can also be used for the control of insect pests

CC through the insertion of a gene encoding an insect toxin into the

CC expression vector which will infect the target pest and produce large

CC quantities of the toxin. Spheroidin and thymidine kinase are

CC nonessential proteins which makes them ideal for the insertion of

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:07:05 ; Search time 23.68 Seconds
(without alignments)
281.799 Million cell updates/sec

Title: US-09-490-700-41

Perfect score: 959

Sequence: 1 FTIPLAEQDCEKFAFTIPAI.....SNLFSILRGSDSLNSKRMLT 182

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	933	97.3	874	1 POLI_HUMAN	P10266 homo sapien
2	525	54.7	870	1 POL_JSRV	P31823 sheep pulmo
3	503	52.5	867	1 POL_SRMV2	P51517 simian retr
4	496	51.7	867	1 POL_SRMV2	P07572 simian maso
5	493	51.4	863	1 POL_IPHA	P04026 hamster int
6	484	50.5	867	1 POL_SRV1	P04025 simian retr
7	474	49.4	888	1 POL_SMRVH	P03364 squirrel mo
8	464	48.4	899	1 POL_SMRVB	P03365 mouse mamma
9	461	48.1	867	1 POL_IPMA	P11368 mouse intra
10	406	42.3	814	1 POL_IPMAI	P12894 mouse intra
11	394	41.1	895	1 POL_RSVP	P03354 rous sarcom
12	281.5	29.4	1036	1 POL_HV2RO	P04584 human immu
13	280.5	29.2	1034	1 POL_HV2CA	P24107 human immu
14	278.5	29.0	1022	1 POL_SIVSP	P19505 simian immu
15	278	29.0	1145	1 POL_ETAVY	P03371 equine infe
16	277.5	28.9	1019	1 POL_SIVS4	P12502 simian immu
17	276.5	28.8	1035	1 POL_HV2N2	P05962 human immu
18	276.5	28.8	1054	1 POL_SIVMK	P05897 simian immu
19	276.5	28.8	1056	1 POL_SIVM1	P05896 simian immu
20	275.5	28.7	1146	1 POL_ETAV9	P11204 equine infe
21	275.5	28.7	1146	1 POL_ETAV9	P32442 equine infe
22	274.5	28.6	1035	1 POL_HV2SB	P12451 human immu
23	273.5	28.5	1073	1 POL_HV2D1	P17757 human immu
24	270.5	28.2	1035	1 POL_HV2KR	Q74120 human immu
25	270.5	28.2	1055	1 POL_HV2ST	P20876 human immu
26	270.5	28.2	1058	1 POL_HV2D2	P15833 human immu
27	270.5	28.2	1142	1 POL_HV2BE	P18096 human immu
28	268	27.9	852	1 POL_BLVAV	P25059 bovine leuk
29	260.5	27.2	1049	1 POL_HV2G1	P18042 human immu
30	255.5	26.6	1047	1 POL_SIVAI1	P27973 simian immu
31	249	26.0	896	1 POL_HV2L1A	P03362 human t-cel
32	246	25.7	852	1 POL_BLVJ	P03361 bovine leuk
33	244.5	25.5	1027	1 POL_SIVC2	P17283 chimpanzee

ALIGNMENTS

RESULT	1
POL1_HUMAN	
ID	POL1_HUMAN STANDARD; PRT; 874 AA.
AC	P10266;
DT	01-MAR-1989 (Rel. 10, Created)
DT	01-MAR-1989 (Rel. 10, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	RETROVIRUS-RELATED POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE
DE	(BC 2.7.7.49); ENDONUCLEASE].
GN	POL.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=87036922; PubMed=3021993;
RA	Ono M., Yasunaga T., Miyata T., Ushikubo H.;
RT	"Nucleotide sequence of human endogenous retrovirus genome related to
RL	J. Virol. 60:589-598(1986).
RL	PIR: D24483; GNHUER.
DR	HSSP: P03366; 1HW.
DR	InterPro: IPR001037; Integrase_C.
DR	InterPro: IPR003308; Integrase_zn.
DR	InterPro: IPR002156; RNaseH.
DR	InterPro: IPR000477; RVTse.
DR	InterPro: IPR001584; Rve.
DR	Pfam: PF00552; Integrase; 1.
DR	Pfam: PF02022; Integrase_zn; 1.
DR	Pfam: PF00075; RNaseH; 1.
DR	Pfam: PF00665; rve; 1.
DR	Pfam: PF00078; rvt; 1.
KW	Hydrolase; Transferase; RNA-directed DNA polymerase; Nuclease;
KW	Endonuclease; Polyprotein.
FT	CHAIN 36 250 REVERSE TRANSCRIPTASE.
FT	CHAIN 585 764 ENDONUCLEASE.
SQ	SEQUENCE 874 AA; 98936 MW; FD985989798018B6 CRC64;

Query Match 97.3%; Score 933; DB 1; Length 874;
Best Local Similarity 97.3%; Pred. No. 2.2e-81;
Matches 177; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY	1	FTIPLAEQDCEKFAFTIPAINNKPEATRFQKVLPOGMLNSPTTCQTFVGRALQPVDRKF	60
DB	128	FTIPLAEQDCEKFAFTIPAINNKPEATRFQKVLPOGMLNSPTTCQTFVGRALQPVREKF	187
QY	61	SDCYIIHYFDILCAETKDKLIDCYTFLPAEVANAGLAIASDKIQSTPFFHYLGMQIEN	120
DB	188	SDCYIIHYFDILCAETKDKLIDCYTFLPAEVANAGLAIASDKIQSTPFFHYLGMQIEN	247
QY	121	RKTKPKQIEIRKDTLKTLDNFQKLGLDINWIRPTLGIPTYAMSLNLSILRGDSDLNSKRM	180
DB	248	RKTKPKQIEIRKDTLKTLDNFQKLGLDINWIRPTLGIPTYAMSLNLSILRGDSDLNSQRI	307

QY 181 LT 182
||
Db 308 LT 309

RESULT 2

POL_JSRV POL_JSRV STANDARD; PRT; 870 AA.
AC P31623;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE (EC 2.7.7.49);
DE ENDONUCLEASE].
GN POL.
OS Sheep pulmonary adenomatosis virus (Jaagsiekte sheep retrovirus)
OS (JSRV).
OC Viruses; Retrovirdae; Retroviridae; Betaretrovirus.
OX NCBI_TaxID=11746;
RN [1]
RP MEDLINE=92333675; PubMed=1629959;
RA York D.F., Vigne R., Verwoerd D.W., Querat G.;
RT "Nucleotide sequence of the jaagsiekte retrovirus, an exogenous and
RT endogenous type D and B retrovirus of sheep and goats.";
RL J. Virol. 66:4930-4939(1992).
CC -|- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -|- MISCELLANEOUS: THIS PROTEIN IS PROBABLY EXPRESSED AS A FUSED
CC GAG-PRO-POL POLYPROTEIN BY A RIBOSOMAL FRAMESHIFTING.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M80216; AAA89182.1; ALT_INIT.
DR PIR; C42740; GNMVJA.
DR HSP; P03366; IHMV.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_Zn.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR000477; RVTse.
DR InterPro; IPR001584; Rve.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF02022; Integrase; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00078; rvt; 1.
KW Transferase; RNA-directed DNA polymerase; Hydrolase; Nuclease;
KW Endonuclease; Polyprotein.
SQ SEQUENCE 870 AA; 99312 MW; CFADAF8D879C033 CRC64;

Query Match 54.7%; Score 525; DB 1; Length 870;
Best Local Similarity 53.3%; Pred. No. 2.1e-42;
Matches 97; Conservative 31; Mismatches 54; Indels 0; Gaps 0;

QY 1 FTPLAEDCEKFAFTIPAINNKEPATRFQWKVLPQGMNSPTTCQTFVGRALQPVDRKF 60
Db YTTPLAEDCEKFAFTIPAINNKEPATRFQWKVLPQGMNSPTTCQTFVGRALQPVDRKF 185
61 SDCVIIHYFDDILCAETKDKLIDCYTFLPAEVANAGLAIASDKIQSTPFGYLGMOIEN 120
Db POLYLVHYMDDILLAHDEHLLYQAFSLKQHLNGLVIADEKIQTFPNYLGFSLYP 245
121 RKIKPQKIEIKDKTLTKLNDPQKLLGDNINWIRPTLGIPTYAMSNLFSILRGSDLSNKR 180
246 RVYNTQLVKLQTDHLTKLNDPQKLLGDNINWIRPTLGIPTYAMSNLFSILRGSDLSNKR 305
QY 181 LT 182

Db 306 LS 307

RESULT 3

POL_SRV2 POL_SRV2 STANDARD; PRT; 867 AA.
AC P51517;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE (EC 2.7.7.49);
DE ENDONUCLEASE].
GN POL.
OS Simian retrovirus SRV-2.
OC Viruses; Retrovirdae; Retroviridae.
OX NCBI_TaxID=39068;
RN [1]
RP MEDLINE=87151131; PubMed=2435057;
RA Thayer R.M., Power M.D., Bryant M.L., Gardner M.B., Barr P.J.,
RA Luciw P.A.;
RT "Sequence relationships of type D retroviruses which cause simian
RT acquired immunodeficiency syndrome.";
RL Virology 157:317-329(1987).
CC -|- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M16605; AAA47562.1;
DR HSP; P04585; IRTI.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_Zn.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR000477; RVTse.
DR InterPro; IPR001584; Rve.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF02022; Integrase; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00078; rvt; 1.
KW Transferase; RNA-directed DNA polymerase; Hydrolase; Nuclease;
KW Endonuclease; Polyprotein.
SQ SEQUENCE 867 AA; 98038 MW; F64227F7365F3659 CRC64;

Query Match 52.5%; Score 503; DB 1; Length 867;
Best Local Similarity 52.2%; Pred. No. 2.7e-40;
Matches 95; Conservative 31; Mismatches 56; Indels 0; Gaps 0;

QY 1 FTPLAEDCEKFAFTIPAINNKEPATRFQWKVLPQGMNSPTTCQTFVGRALQPVDRKF 60
Db YTTPLAEDCEKFAFTIPAINNKEPATRFQWKVLPQGMNSPTTCQTFVGRALQPVDRKF 185
61 SDCVIIHYFDDILCAETKDKLIDCYTFLPAEVANAGLAIASDKIQSTPFGYLGMOIEN 120
Db AQMYIIHYMDDILLAHDEHLLYQAFSLKQHLNGLVIADEKIQTFPNYLGFSLYP 245
121 RKIKPQKIEIKDKTLTKLNDPQKLLGDNINWIRPTLGIPTYAMSNLFSILRGSDLSNKR 180
246 PKITNQKAVIRDRKQLTKLNDPQKLLGDNINWIRPTLGIPTYAMSNLFSILRGSDLSNKR 305
QY 181 LT 182
Db 306 LS 307

RP SEQUENCE FROM N.A.
RX MEDLINE=86151668; PubMed=3006247;
RA Power M.D., Marx P.A., Bryant M.L., Gardner M.B., Barr P.J.,
RA Luciw P.A.;
RT "Nucleotide sequence of SRV-1, a type D simian acquired immune
RT deficiency syndrome retrovirus.";
RL Science 231:1567-1572(1986).
CC -1- PPM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11841; AAA47732.1; -;
DR PIR; A03963; GNLJSA.
DR HSSP; P04585; LRTI.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_zn.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR000477; Rvise.
DR InterPro; IPR001584; Rve.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF02022; Integrase_zn; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; Ives; 1.
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; Ives; 1.
DR Pfam; PF00078; rvt; 1.
KW Transferase; RNA-directed DNA polymerase; Hydrolase; Nuclease;
KW Endonuclease; Polyprotein.
SQ SEQUENCE 867 AA; 98361 MW; ABB06A0B770A6473 CRC64;

Query Match 50.5%; Score 484; DB 1; Length 867;
Best Local Similarity 52.2%; Pred. No. 1.7e-38;
Matches 95; Conservative 27; Mismatches 60; Indels 0; Gaps 0;

Qy 1 FTIPLAEDCEKEAFTTIPAINNKEPATRFQWKLPOGMLNSPTICQTFVGRALOPVRDKF 60
Db 126 FSIPLHSDQRKAFSLFNFKEPMQRFQWKLPOGMLNSPTICQTFVGRALOPVRDKF 185
Qy 61 SDCVIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLAIASDKIQTSTPFFYLGQMQIEN 120
Db 186 KQVLIHYMDILLIAGKDGQVQVLCDFQDLKQELTLAGLHIAPEKIQLPDPTYLGFELNG 245
Qy 121 RKIKPQKIEIKRDKLTLDNQKLLGNDINWIRPTLGIPTTYAMSNLFSILRGDSDLNSKRM 180
Db 246 PKITNQKAVIRKDKLTLDNQKLLGNDINWIRPTLGIPTTYAMSNLFSILRGDSDLNSKRM 305
Qy 181 LT 182
Db 306 LS 307

RESULT 7

POL_SMRVH
ID POL_SMRVH STANDARD; PRT; 888 AA.
AC P03364;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE (EC 2.7.7.49);
DE ENDONUCLEASE].
GN POL.
OS Squirrel monkey retrovirus (SMRV-H) (SMRV-HLB).
OC Viruses; Retroviral viruses; Retroviridae; Betaretrovirus.
OX NCBI_Taxid=11856;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89073750; PubMed=3201749;
RA Oda T., Ikeda S., Watanabe S., Hatsushika M., Akiyama K.,

RA Mitsunobu F.;
RT "Molecular cloning, complete nucleotide sequence, and gene structure
RT of the provirus genome of a retrovirus produced in a human
RT lymphoblastoid cell line.";
RL Virology 167:468-476(1988).
RN [2]
RP SEQUENCE OF 595-774 FROM N.A.
RX MEDLINE=84097535; PubMed=6197754;
RA Chiu I.-M., Callahan R., Tronick S.R., Schlom J., Aaronson S.A.;
RT "Major pol. gene progenitors in the evolution of oncoviruses.";
RL Science 223:364-370(1984).
CC -1- PPM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K01706; AAA46815.1; -;
DR EMBL; M23385; AAA68453.1; ALT_INIT.
DR PIR; C31827; GNLJHD.
DR PIR; A05072; A05072.
DR HSSP; P03366; LHMV.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_zn.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR000477; Rvise.
DR InterPro; IPR001584; Rve.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF02022; Integrase_zn; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; Ives; 1.
DR Pfam; PF00078; rvt; 1.
KW Transferase; RNA-directed DNA polymerase; Hydrolase; Nuclease;
KW Endonuclease; Polyprotein.
SQ SEQUENCE 888 AA; 99165 MW; B6B2CD09C651B98E CRC64;

Query Match 49.4%; Score 474; DB 1; Length 888;
Best Local Similarity 49.5%; Pred. No. 1.6e-37;
Matches 90; Conservative 35; Mismatches 57; Indels 0; Gaps 0;

Qy 1 FTIPLAEDCEKEAFTTIPAINNKEPATRFQWKLPOGMLNSPTICQTFVGRALOPVRDKF 60
Db 123 FTIPLHEDRPYFAFSPQINFQSPMPRIQWKLPOGMLNSPTICQTFVGRALOPVRDKF 182
Qy 61 SDCVIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLAIASDKIQTSTPFFYLGQMQIEN 120
Db 183 PEAVILHYMDILLACDSAAKACAYAHIIISCLTSYGLKIAPDKVQVSEPFYLGFLHH 242
Qy 121 RKIKPQKIEIKRDKLTLDNQKLLGNDINWIRPTLGIPTTYAMSNLFSILRGDSDLNSKRM 180
Db 243 QQVFTPRVCLKTDLHLKLTLDNQKLLGNDINWIRPTLGIPTTYAMSNLFSILRGDSDLNSKRM 302
Qy 181 LT 182
Db 303 LT 304

RESULT 8

POL_MMTVB
ID POL_MMTVB STANDARD; PRT; 899 AA.
AC P03365;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE (EC 2.7.7.49);
DE ENDONUCLEASE].
GN POL.
OS Mouse mammary tumor virus (strain BR6).

OC Viruses; Retroid viruses; Retroviridae; Betaretrovirus.
 OX NCBI_TaxID=11758;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87112944; PubMed=3027377;
 RA Moore R., Dixon M., Smith R., Peters G., Dickson C.;
 RT "Complete nucleotide sequence of a milk-transmitted mouse mammary
 tumor virus: two frameshift suppression events are required for
 translation of gag and pol.";
 RL J. Virol. 61:480-490(1987).
 RN [2]
 RP SEQUENCE OF 578-757 FROM N.A.
 RX MEDLINE=84097535; PubMed=6197754;
 RA Chiu I.-M., Callahan R., Tronick S.R., Schlom J., Aaronson S.A.;
 RT "Major pol gene progenitors in the evolution of oncoviruses.";
 RL Science 223:364-370(1984).
 CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: K01707; AAA46540.1; -;
 DR EMBL: M15122; AAA46542.1; -;
 DR PIR: C26795; GNKVM.4.
 DR HSSP: P03366; IHMV.
 DR InterPro: IPR001037; Integrase_C.
 DR InterPro: IPR003308; Integrase_zn.
 DR InterPro: IPR002156; RNaseH.
 DR InterPro: IPR000477; RVTse.
 DR InterPro: IPR001584; Rve.
 DR Pfam: PF00552; Integrase; 1.
 DR Pfam: PF02022; Integrase_zn; 1.
 DR Pfam: PF00075; rnaseH; 1.
 DR Pfam: PF00665; rve; 1.
 DR Pfam: PF00078; rvt; 1.
 KW Transferase: RNA-directed DNA polymerase; Hydrolase; Nuclease;
 KW Endonuclease; Polyprotein.
 FT CONFLICT 625 625 H -> S (IN REF. 2).
 SQ SEQUENCE 899 AA; 102177 MW; 0FC151C3EBE3C417 CRC64;

Query Match 48.4%; Score 464; DB 1; Length 899;
 Best Local Similarity 50.0%; Pred. No. 1.5e-36;
 Matches 91; Conservative 30; Mismatches 61; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCKEFAFTIPAINNKEPATRFQKVLPOGMLNSPTICQTFVGRALQPVDRKF 60
 Db 120 FNILHPEDCKRAFSVPSNFRFYQRFQKVLPOGKNSPTLCQKFVDKAILTVRDKY 179
 QY 61 SDCYIIHYFDILCAAEKDKLIDCYTFLPAEVANAGLAISDKIQSTSTPFHYLGMQIEN 120
 Db 180 QDSYIVHYMDILLAPSRSDVDEILTSMTQALNKHGLVSTEKIQYDNLKYLGLTHIQG 239
 QY 121 RKIKPKIEIRKDTLTKLNDFOKLGDINWIRPTLGIPTTYAMNSILRGDSDLNSKRM 180
 Db 240 DSVSYQKLIQRTDKLRTLNDFOKLGDINWIRPFLKLTGTGELKPLFEILGDSNPISIRK 299
 QY 181 LT 182
 Db 300 LT 301

RESULT 9
 POL_IPMAI
 ID POL_IPMAI STANDARD; PRT; 867 AA.
 AC P11368;
 DT 01-JUL-1989 (Rel. 11, Created)
 CC -1- MISCELLANEOUS: THIS PARTICLE IS A DEFECTIVE RETROVIRUS.

DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PUTATIVE POL POLYPROTEIN [CONTAINS: ENDONUCLEASE; REVERSE
 GN TRANSCRIPTASE (EC 2.7.7.49)].
 OS POL.
 OC Mouse intracisternal a-particle (IAP-MIA14).
 OS Viruses; Retroid viruses; Retroviridae; Intracisternal A-particles.
 RN NCBI_TaxID=11753;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87311859; PubMed=3041022;
 RA Mietz J.A., Grossman Z., Lueders K.K., Kuff E.L.;
 RT "Nucleotide sequence of a complete mouse intracisternal A-particle
 RT genome: relationship to known aspects of particle assembly and
 RT function.";
 RL J. Virol. 61:3020-3029(1987).
 CC -1- MISCELLANEOUS: THIS PARTICLE IS A DEFECTIVE RETROVIRUS.
 DR PIR: B26787; GNMSIA.
 DR HSSP: P04585; IKLM.
 DR InterPro: IPR001037; Integrase_C.
 DR InterPro: IPR003308; Integrase_zn.
 DR InterPro: IPR002156; RNaseH.
 DR InterPro: IPR000477; RVTse.
 DR InterPro: IPR001584; Rve.
 DR Pfam: PF00552; Integrase; 1.
 DR Pfam: PF02022; Integrase_zn; 1.
 DR Pfam: PF00075; rnaseH; 1.
 DR Pfam: PF00665; rve; 1.
 DR Pfam: PF00078; rvt; 1.
 KW Transferase: RNA-directed DNA polymerase; Hydrolase; Nuclease;
 KW Endonuclease; Polyprotein.
 SQ SEQUENCE 867 AA; 97778 MW; 7394B47ED63235B6 CRC64;

Query Match 48.1%; Score 461; DB 1; Length 867;
 Best Local Similarity 51.1%; Pred. No. 2.7e-36;
 Matches 93; Conservative 28; Mismatches 61; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCKEFAFTIPAINNKEPATRFQKVLPOGMLNSPTICQTFVGRALQPVDRKF 60
 Db 123 FSIPLCPRDRPRAFTIPISINDEPNRYQKVLPOGMSNPTWCQLYVQEALLPVREQF 182
 QY 61 SDCYIIHYFDILCAAEKDKLIDCYTFLPAEVANAGLAISDKIQSTSTPFHYLGMQIEN 120
 Db 183 PSILLLYMDILLAPSRSDVDEILTSMTQALNKHGLVSTEKIQYDNLKYLGLTHIQG 242
 QY 121 RKIKPKIEIRKDTLTKLNDFOKLGDINWIRPTLGIPTTYAMNSILRGDSDLNSKRM 180
 Db 243 DKIVPKVEIRDRDLHTLNNFQKLLGDINWIRPFLKIPSAELRPLFWYLEGDPHISPT 302
 QY 181 LT 182
 Db 303 LT 304

RESULT 10
 POL_IPMAI
 ID POL_IPMAI STANDARD; PRT; 814 AA.
 AC P12894;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE PROBABLE POL POLYPROTEIN.
 OS Mouse intracisternal a-particle (IAP-IL3).
 OC Viruses; Retroid viruses; Retroviridae; Intracisternal A-particles.
 OX NCBI_TaxID=11754;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86286596; PubMed=3016667;
 RA Ymer S., Tucker W.Q.J., Campbell H.D., Young I.G.;
 RT "Nucleotide sequence of the intracisternal A-particle genome inserted
 RT 5' to the interleukin-3 gene of the leukemia cell line WEHI-38.";
 RL Nucleic Acids Res. 14:5901-5918(1986).
 CC -1- MISCELLANEOUS: THIS PARTICLE IS A DEFECTIVE RETROVIRUS.

RT "Crystal structure at 1.9-A resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally available inhibitor of the HIV proteases.";

RL J. Biol. Chem. 269:26344-26348(1994).

RN [4]

RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 86-184 IN COMPLEX WITH INHIB.

RX MEDLINE=95338600; PubMed=7613867;

RA Priestle J.P., Fassler A., Rosel J., Tintelnot-Blomley M., Strop P., Grutter M.G.;

RT "Comparative analysis of the X-ray structures of HIV-1 and HIV-2 proteases in complex with C6P 53820, a novel pseudosymmetric inhibitor.";

RL Structure 3:381-389(1995).

RN [5]

RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 86-184 IN COMPLEX WITH INHIB.

RX MEDLINE=97359919; PubMed=9216835;

RA Beaulieu P.L., Wernic D., Abraham A., Anderson P.C., Bogri T., Bousquet Y., Croteau G., Guse I., Lamarre D., Liard F., Paris W., Thibeault D., Pav S., Tong L.;

RT "Potent HIV protease inhibitors containing a novel (hydroxyethyl)amide isostere.";

RL J. Med. Chem. 40:2164-2176(1997).

CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE DETERMINED.

CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO KNOWN AS THE RETROPEPSIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M15390; -; NOT_ANNOTATED.CDS.

DR EMBL; X05291; -; NOT_ANNOTATED.CDS.

DR EMBL; M15390; AAB00764.1; ALT_INIT.

DR PIR; B26262; GNJG2.

DR PDB; 1PHV; 15-OCT-94.

DR PDB; 2PHV; 15-OCT-94.

DR PDB; 2MIP; 15-APR-92.

DR PDB; 2HPE; 15-OCT-94.

DR PDB; 2HPE; 15-OCT-94.

DR PDB; 1HII; 10-JUL-95.

DR PDB; 1HSN; 03-APR-96.

DR PDB; 1HSI; 03-APR-96.

DR PDB; 1IDA; 26-JAN-95.

DR PDB; 1JLD; 26-JAN-95.

DR PDB; 3UPJ; 14-OCT-96.

DR PDB; 4UPJ; 14-OCT-96.

DR PDB; 5UPJ; 21-APR-97.

DR PDB; 6UPJ; 21-APR-97.

DR HIV; M15390; POL\$2ROD.

DR MEROPS; A02.002; -.

DR InterPro; IPR001995; Asp_prot_retrov.

DR InterPro; IPR001969; Asp_protease.

DR InterPro; IPR001037; Integrase_C.

DR InterPro; IPR003308; Integrase_zn.

DR InterPro; IPR002156; RNaseH.

DR InterPro; IPR000477; RVse.

DR InterPro; IPR001584; Rve.

DR Pfam; PF02022; Integrase_1.

DR Pfam; PF00075; rnaseh; 1.

DR Pfam; PF00665; rve; 1.

DR Pfam; PF00077; rvp; 1.

DR Pfam; PF00078; rvt; 1.

DR PROSITE; PS00141; ASP_PROTEASE; 1.

DR PROSITE; PS0175; ASP_PROT_RETROV; 1.

KW AIDS; Polypeptin; Hydrolase; Aspartyl protease; Endonuclease; Nuclease; Transferase; RNA-directed DNA polymerase; 3D-structure.

FT CHAIN 86 184 PROTEASE.

FT ACT_SITE 110 110 BY SIMILARITY.

FT STRAND 87 89

FT STRAND 95 100

FT TURN 101 102

FT STRAND 103 109

FT TURN 111 112

FT STRAND 117 118

FT STRAND 128 134

FT STRAND 138 151

FT TURN 152 153

FT STRAND 154 162

FT STRAND 169 170

FT HELIX 172 178

FT TURN 179 179

FT STRAND 181 183

SQ SEQUENCE 1036 AA; 117080 MW; 5224E354B1DCC83B CRC64;

Query Match 29.4%; Score 281.5; DB 1; Length 1036;

Best Local Similarity 37.2%; Pred. No. 4.9e-19;

Matches 67; Conservative 30; Mismatches 78; Indels 5; Gaps 3;

QY 1 FTPLAEQDCEKEAFTIPAINNKPEATRFQWKVLPGQMLNSPTICQTFVGRALQPVDRKF 60

DB 300 FSIPLHEDFRPYTAFTLPSYNNAPGKRYTKVLPQGWKSPAIQFTMRQVLEPFRKAN 359

QY 61 SDCVIITHYFDILCAAEWKDLIDCYTFLPAEVANA-GLAIASDKIOTSTPHYLGMOIE 119

DB 360 KDVIITQYMDILASDRTDLEHDRVVLQKELLGLGFTPDDEKFKODPPHYMMGYELW 419

QY 120 NRKIKPKQIEIRKDTLKTLDNFQKLGIDINWIRPTL-GIPTYAMSNLFSILRGSDSLNSK 178

DB 420 PTKWKLQKIQLPKEIWTVDNIQKLVGLNAAQLYPGIKT---KHLCLRLIRCKMTLTEE 476

RESULT 13

POL_HV2CA STANDARD; PRT; 1034 AA.

AC P24107;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16); DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].

OS Human immunodeficiency virus type 2 (isolate CAM2) (HIV-2).

OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11715;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91170959; PubMed=2005437;

RA Tristram M., Hill F., Karpas A.;

RT "Nucleotide sequence of a Guinea-Bissau-derived human immunodeficiency virus type 2 proviral clone (HIV-2CAM2).";

RL J. Gen. Virol. 72:721-724(1991).

CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE DETERMINED.

CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR TGA MAY OCCUR BETWEEN 564-ILE AND 565-GLY.

CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO KNOWN AS THE RETROPEPSIN FAMILY.

CC PIR; B38475; GNJCA.

DR HSSP; P04584; 1JLD.

DR MEROPS; A02.002; -.

DR InterPro; IPR001995; Asp_prot_retrov.

DR InterPro; IPR001969; Asp_protease.

DR InterPro; IPR001037; Integrase_C.

DR InterPro; IPR003308; Integrase_zn.

DR InterPro; IPR002156; RNaseH.

DR InterPro; IPR000477; RVse.

DR InterPro; IPR001584; Rve.

DR Pfam; PF00552; Integrase; 1.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 16:58:15 ; Search time 39.41 Seconds
(without alignments)
351.783 Million cell updates/sec

Title: US-09-490-700-41

Perfect score: 959

Sequence: 1 FTIPLAEQDCEKFAFTIPAI.....SNLFSILRGSDSLNSKRMLT 182

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	933	97.3	874	4 GNMUJ	retrovirus-related
2	525	54.7	870	1 GNMVJA	pol polyprotein -
3	496	51.7	867	1 GNLJMP	pol polyprotein (c
4	493	51.4	863	1 GNMHII	retrovirus-related
5	490	51.1	752	2 D40899	pol polyprotein -
6	484	50.5	867	1 GNLJSA	pol polyprotein -
7	474	49.4	888	1 GNLJHD	pol polyprotein -
8	464	48.4	899	1 GNMVMM	pol polyprotein -
9	461	48.1	867	1 GNMVIA	retrovirus-related
10	406	42.3	814	1 GNMVIP	retrovirus-related
11	404	42.1	365	3 JC7527	nuclear retroviral
12	395	41.2	895	2 S35429	pol polyprotein -
13	394	41.1	895	1 GNFVIR	pol polyprotein -
14	381	39.7	896	2 G48613	pol polyprotein -
15	381	39.7	896	2 E48613	pol polyprotein -
16	381	39.7	1603	2 A48613	gag/pol polyprotei
17	346.5	36.1	843	2 S33123	pol polyprotein -
18	285.5	29.8	1019	2 T11560	pol polyprotein -
19	282.5	29.5	656	2 S30484	pol polyprotein -
20	281.5	29.4	1036	1 GNLJG2	pol polyprotein -
21	280.5	29.2	1034	1 GNLJCA	pol polyprotein -
22	278.5	29.0	1055	2 S53092	pol polyprotein -
23	278	29.0	1145	1 GNLJEV	pol polyprotein -
24	276.5	28.8	656	2 S30483	pol polyprotein -
25	276.5	28.8	1054	1 GNLJG5	pol polyprotein -
26	276.5	28.8	1056	1 GNLJG3	pol polyprotein -
27	275.5	28.7	1146	1 GNLJEV	pol polyprotein (c
28	275.5	28.7	1146	1 GNLJ22	pol polyprotein -
29	273.5	28.5	1032	2 S12153	pol polyprotein -

30	271	28.3	852	2 S29358	pol protein - bovi
31	270.5	28.2	1055	1 GNLJST	pol polyprotein -
32	270.5	28.2	1058	2 S08436	pol polyprotein -
33	268	27.9	852	1 GNLJGA	pol polyprotein -
34	260.5	27.2	1035	1 GNLJGG	pol polyprotein -
35	251.5	26.2	1039	2 S46347	pol polyprotein -
36	249	26.0	896	1 GNLJGH	pol polyprotein -
37	246	25.7	852	1 GNLJGB	pol polyprotein -
38	244.5	25.5	1027	1 GNLJSI	pol polyprotein -
39	241	25.1	982	1 GNLJH2	pol polyprotein -
40	240	25.0	896	1 GNLJCN	pol polyprotein -
41	237.5	24.8	559	2 B47175	reverse transcript
42	237.5	24.8	1003	1 GNMVLV	pol polyprotein -
43	237.5	24.8	1012	1 GNMVLV	pol polyprotein -
44	236.5	24.7	559	2 A47175	reverse transcript
45	235.5	24.6	1053	1 GNLJBT	pol polyprotein -

ALIGNMENTS

RESULT 1

GNHUE

retrovirus-related pol polyprotein pseudogene - human

N:Contains: endonuclease (EC 3.1.-.-); RNA-directed DNA polymerase (EC 2.7.7.49)

C:Species: Homo sapiens (man)

C>Date: 28-Dec-1987 #sequence_revision 04-Jan-1996 #text_change 14-May-1999

C:Accession: D24483

R:Ono, M.; Yasunaga, T.; Miyata, T.; Ushikubo, H.

J. Virol. 60, 589-598, 1986.

A:Title: Nucleotide sequence of human endogenous retrovirus genome related to the mou

A:Reference number: A93023; MUID:87036922

A:Accession: D24483

A>Status: conceptual translation of pseudogene

A:Molecule type: DNA

A:Residues: 1-874 <ONO>

A:Cross-references: GB:M14123; NID:g182227

C:Genetics:

A:Gene: pol

C:Keywords: endonuclease; hydrolase; nucleotidyltransferase; polyprotein; pseudogene;

F:36-250/Domain: RNA-directed DNA polymerase <REV>

F:585-764/Domain: endonuclease <END>

Query Match 97.3%; Score 933; DB 4; Length 874;
Best Local Similarity 97.3%; Pred No. 4.5e-79;
Matches 177; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGMNLSPTICQTFVGRALQPVREKF 60

|||||

Db 128 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGMNLSPTICQTFVGRALQPVREKF 187

Oy 61 SDCYIIHYFDDILCAATKDKLIDCYTFLEPAEVANAGLATASDKIQTSTPHYLGMQIEN 120

|||||

Db 188 SDCVIIHYIDDILCAATKDKLIDCYTFLEAEVANAGLATASDKIQTSTPHYLGMQIEN 247

Oy 121 RKIKPQKIEIRKDTLKTNDPQKLLGDIINIRPTLGIPTYAMSNLFSILRGSDSLNSKRM 180

|||||

Db 248 RKIKPQKIEIRKDTLKTNDPQKLLGDIINIRPTLGIPTYAMSNLFSILRGSDSLNSQRI 307

Oy 181 LT 182

||

Db 308 LT 309

RESULT 2

GNMVJA

pol polyprotein - sheep pulmonary adenomatosis virus

N:Contains: endonuclease (EC 3.1.-.-); RNA-directed DNA polymerase (EC 2.7.7.49)

C:Species: sheep pulmonary adenomatosis virus

C>Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 08-Apr-1994

C:Accession: C42740

R:York, D.F.; Vigne, R.; Verwoerd, D.W.; Querat, G.

J. Virol. 66, 4930-4939, 1992
A:Title: Nucleotide sequence of the Jaagsiekte retrovirus, an exogenous and endogenous t
A:Reference number: A42740; MUID:92333675
A:Accession: C42740
A:Molecule type: genomic RNA
A:Residues: 1-870 <YOR>
A:Cross-references: GB:M80216
C:Comment: This protein is likely to be expressed as a gag-pol polyprotein.
C:Comment: The precise boundary between RNA-directed DNA polymerase and endonuclease has
C:Genetics:
A:Gene: pol
A:Start codon: UCA
C:Superfamily: pol polyprotein
C:Keywords: endonuclease; hydrolase; nucleotidyltransferase; polyprotein; reverse transcr

Query Match 54.78; Score 525; DB 1; Length 870;
Best Local Similarity 53.3%; Pred. No. 5.6e-41;
Matches 97; Conservative 31; Mismatches 54; Indels 0; Gaps 0;
QY 1 FTIPLAEDCEKFAFTIPAINNKEPATRFQWKVLPQGMNSPTTCQTFVGRALQPVDRKF 60
DB 126 YTIPLAQDCKRFAFSLPSVNFEPQWRYQWRVLPQGMNSPTTCQKQFVATAIPVRQRF 185
QY 61 SDCVIIHYFDDILCAAEFTKDLIDCYTFPLPAEVANAGLAIASDKIQSTPTPHYLGMOIEN 120
DB 186 POLYLVHYMDDILLAHNTDEHLLYQAFSLKQHLGLSLGLVIADEKIQTHFPYNYLGFSLYP 245
QY 121 RKIRPKQIEIRKDKLTLDNQKLLGDNINWIRPTLGIPTIYAMSNLFSILRGDSDLSNKR 180
DB 246 RYNTQVLKLTQDHLKLTLDNQKLLGDNINWIRPKLPTTYTLQPLFLDLKGDSDPASPT 305
QY 181 LT 182
DB 306 LS 307

RESULT 3
GNLJMP
pol polyprotein (clone 6A) - Mason-Pfizer monkey virus
N:Name: endonuclease (EC 3.1.1.-); RNA-directed DNA polymerase (EC 2.7.7.49)
C:Species: Mason-Pfizer monkey virus
C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999
C:Accession: C25839
R:Sonigo, P.; Barker, C.; Hunter, E.; Wain-Hobson, S.
Cell 45, 375-385, 1986
A:Title: Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type x
A:Reference number: A90878; MUID:86189951
A:Accession: C25839
A:Molecule type: DNA
A:Residues: 1-867 <SON>
A:Cross-references: GB:M12349; NID:g334702; PID:AAA47711.1; PID:g334704
C:Comment: The pol polyprotein contains reverse transcriptase and endonuclease; however,
C:Genetics:
A:Gene: pol
C:Superfamily: pol polyprotein
C:Keywords: AIDS; endonuclease; hydrolase; immunodeficiency; nucleotidyltransferase; pol

Query Match 51.78; Score 496; DB 1; Length 867;
Best Local Similarity 52.7%; Pred. No. 2.8e-38;
Matches 96; Conservative 27; Mismatches 59; Indels 0; Gaps 0;
QY 1 FTIPLAEDCEKFAFTIPAINNKEPATRFQWKVLPQGMNSPTTCQTFVGRALQPVDRKF 60
DB 126 FSIPLHPSDQKRFAPSLSTNFKPEPMQRFQWKVLPQGMNSPTTCQKQVATAIHKVRHAW 185
QY 61 SDCVIIHYFDDILCAAEFTKDLIDCYTFPLPAEVANAGLAIASDKIQSTPTPHYLGMOIEN 120
DB 186 KMVYIHYMDDILLIAGDQGVQVLCFQDLKQELTAAGLHIAPEKVKVQLQDPYTYLGFELNG 245
QY 121 RKIRPKQIEIRKDKLTLDNQKLLGDNINWIRPTLGIPTIYAMSNLFSILRGDSDLSNKR 180
DB 181 LT 182
DB 306 LS 307

Db 246 PKINQXAVIRKDKLTLDNQKLLGDNINWIRPKLTTGDLKPLFDLTGKGDSPNSHRS 305
QY 181 LT 182
DB 306 LS 307

RESULT 4
GNHVIIH
retrovirus-related pol polyprotein - golden hamster intracisternal A-particle H18
N:Alternate names: reverse transcriptase
C:Species: golden hamster intracisternal A-particle H18
A:Note: host Mesocricetus auratus (golden hamster)
C:Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 30-Sep-1993
C:Accession: A03964
R:Ono, M.; Toh, H.; Miyata, T.; Awaya, T.
J. Virol. 55, 387-394, 1985
A:Title: Nucleotide sequence of the Syrian hamster intracisternal A-particle gene: cl
A:Reference number: A93012; MUID:85264989
A:Accession: A03964
A:Molecule type: DNA
A:Residues: 1-863 <ONO>
C:Comment: Readthrough of three terminators may occur: TAA between codons ATT for 660
59-Pro and ATT for 860-Ile.
C:Genetics:
A:Gene: pol
A:Introns: 315/3
C:Superfamily: pol polyprotein
C:Keywords: polyprotein; reverse transcriptase

Query Match 51.4%; Score 493; DB 1; Length 863;
Best Local Similarity 53.3%; Pred. No. 5.4e-38;
Matches 97; Conservative 29; Mismatches 56; Indels 0; Gaps 0;
QY 1 FTIPLAEDCEKFAFTIPAINNKEPATRFQWKVLPQGMNSPTTCQTFVGRALQPVDRKF 60
DB 112 FSIPLYPRDRPRFAFTIPSLNHMEPDKRFQWKVLPQGMNSPTTCQLYVQVLEPIRKQF 171
QY 61 SDCVIIHYFDDILCAAEFTKDLIDCYTFPLPAEVANAGLAIASDKIQSTPTPHYLGMOIEN 120
DB 172 TSLIVHYMDDILLCHKELDVQKAPMLVAELQWGLEIAEKVQIADTGLFGSKITP 231
QY 121 RKIRPKQIEIRKDKLTLDNQKLLGDNINWIRPTLGIPTIYAMSNLFSILRGDSDLSNKR 180
DB 232 KNIVPQKIEIRKDKLTLDNQKLLGDNINWIRPKLIPSAOLKPLFDLLEGEPIHSSPK 291
QY 181 LT 182
DB 292 FT 293

RESULT 5
D40899
pol polyprotein - Chinese hamster intracisternal A-particle CH1AP34
C:Species: Chinese hamster intracisternal A-particle CH1AP34
C:Date: 06-Mar-1992 #sequence_revision 06-Mar-1992 #text_change 07-Feb-1997
C:Accession: D40899
R:Dorner, A.J.; Bonneville, F.; Kriz, R.; Kelleher, K.; Bean, K.; Kaufman, R.J.
J. Virol. 65, 4713-4719, 1991
A:Title: Molecular cloning and characterization of a complete Chinese hamster proviru
A:Reference number: A40899; MUID:91333012
A:Accession: D40899
A:Molecule type: DNA
A:Residues: 1-752 <DOR>
A:Cross-references: GB:M73970
C:Superfamily: pol polyprotein
C:Keywords: polyprotein

Query Match 51.1%; Score 490; DB 2; Length 752;
Best Local Similarity 53.3%; Pred. No. 8.6e-38;


```
RESULT 9
GNMSTA
retrovirus-related pol polyprotein - mouse intracisternal A-particle MIA14
C:Species: mouse intracisternal A-particle MIA14
A:Note: host Mus musculus (house mouse)
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 30-Jun-1993
C:Accession: B26787
R:Metz, J.A.; Grossman, Z.; Lueders, K.K.; Kuff, E.L.
J. Virol. 61, 3020-3029, 1987
A:Title: Nucleotide sequence of a complete mouse intracisternal A-particle genome: related
A:Reference number: A93027; MUID:87311859
A:Accession: B26787
A:Molecule type: DNA
A:Residues: 1-867 <MIE>
A:Note: the authors translated the codon TGG for residue 64 as Asp, AAC for residue 92 a
C:Comment: The DNA sequence was obtained from GenBank, release 55.0.
C:Genetics:
C:Superfamily: pol polyprotein
C:Keywords: polyprotein; reverse transcriptase

Query Match 48.1%; Score 461; DB 1; Length 867;
Best Local Similarity 51.1%; Pred. No. 5.3e-35;
Matches 93; Conservative 28; Mismatches 61; Indels 0; Gaps 0;

Qy 1 FTIPLAEOCEKFAFTIPAINNKEPATFQKVLPGQMLNSPTTCQTFVGRALQPVDRKF 60
Db 123 FSIPLCPDRFRFAFTIPINSDEPNRYQKVLPGQMSNPTWCQLYVQKALLPVRFQF 182
Qy 61 SDCVIIHYFDDILCAAEKDKLIDCYTFPLPAEVANAGLAIASDKIQSTPTPHYLGMQIEN 120
Db 183 PSLILLYMDILLCHKELTKAYPELLTTSQWGLQIATEKVSQSDTQGLGVSVP 242
Qy 121 RKIRPQKIEIRKDLTKLNDQKLLGDNINWIRPLGIPITYAMSNLFSILRGSDSLNSKRM 180
Db 243 DKIVPQKVEIRRDHLTLNNFKLGDINWIRPLFKIPSAELRPLFWLEGDPHISSPT 302
Qy 181 LT 182
Db 303 LT 304

RESULT 10
GNMSTP
retrovirus-related pol polyprotein - mouse intracisternal A-particle MIA13
C:Species: mouse intracisternal A-particle MIA13
A:Note: host Mus musculus (house mouse)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: A23597
R:Ymer, S.; Tucker, W.O.J.; Campbell, H.D.; Young, I.G.
Nucleic Acids Res. 14, 5901-5918, 1986
A:Title: Nucleotide sequence of the intracisternal A-particle genome inserted, 5' to the
A:Reference number: A23597; MUID:86286596
A:Accession: A23597
A:Molecule type: DNA
A:Residues: 1-814 <YME>
A:Cross-references: GB:X04120; NID:g51502; PIDN:CAA27732.1; PID:g51503
C:Comment: This particle is a defective retrovirus.
C:Superfamily: pol polyprotein
C:Keywords: polyprotein; reverse transcriptase

Query Match 42.3%; Score 406; DB 1; Length 814;
Best Local Similarity 51.6%; Pred. No. 6.7e-30;
Matches 82; Conservative 24; Mismatches 53; Indels 0; Gaps 0;

Qy 24 EPATRFQKVLPGQMLNSPTTCQTFVGRALQPVDRKFSDCYIIHYFDDILCAAEKDKLI 83
Db 65 KPDKRYQKVLPGQMSNPTWCQLYVQKALLPVRFQPSLLILLYMDDILLCHKDLTMQ 124
Qy 84 DCYTFPLPAEVANAGLAIASDKIQSTPTPHYLGMQIENRKIPQKIEIRKDKLTNDQK 143
```

```
Db 125 KAYPFLKLTLSQWGLQIATEKVSQSDTQGLGVSVPDKIVPQKVEIRRDHLTLNDQK 184
Qy 144 LLGDNINWIRPLGIPITYAMSNLFSILRGSDSLNSKRMILT 182
Db 185 LLGDNINWIRPLFKIPSAELRPLFSILEGDPHISSPTUT 223

RESULT 11
JC7527
nuclear retroviral polymerase-like protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Mar-2001 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
C:Accession: JC7527
R:Graham, K.M.; Ko, C.; Park, K.; Sarge, K.; Park-Sarge, O.K.
Biochem. Biophys. Res. Commun. 278, 48-57, 2000
A:Title: Expression of an intracisternal A-particle-like element in rat ovary.
A:Reference number: JC7527; MUID:20525406
A:Contents: Ovary, granulosa cells
A:Accession: JC7527
A:Molecule type: mRNA
A:Residues: 1-365 <GRA>
A:Cross-references: GB:AA964260
A:Comment: This protein is involved in cell proliferation, differentiation, folliculo
C:Genetics:
A:Gene: iap-le
C:Keywords: ovary; transformation

Query Match 42.1%; Score 404; DB 3; Length 365;
Best Local Similarity 50.0%; Pred. No. 3.7e-30;
Matches 80; Conservative 24; Mismatches 56; Indels 0; Gaps 0;

Qy 23 KEPATRFQKVLPGQMLNSPTTCQTFVGRALQPVDRKFSDCYIIHYFDDILCAAEKDKL 82
Db 2 KPDKRYQKVLPGQMSNPTWCQLYVQKALLPVRFQKRIIHFMDILLSAKDCSTL 61
Qy 83 IDCYTFPLPAEVANAGLAIASDKIQSTPTPHYLGMQIENRKIPQKIEIRKDKLTNDQ 142
Db 62 ETAYAEVIKLTLESQNFIAPEKQVQKGEYLGAKITPHNVSPQKIEIRKDKLTNDQ 121
Qy 143 KLLGDNINWIRPLGIPITYAMSNLFSILRGSDSLNSKRMILT 182
Db 122 KFMGSINWIRPYINMPNADLQPLYLEILKGDLSQLTSPRLT 161

RESULT 12
S35429
pol polyprotein - avian leukosis virus
C:Species: avian leukosis virus, ALV
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 26-Aug-1999
C:Accession: S35429; S35435
R:Bieth, E.; Darlix, J.L.
Nucleic Acids Res. 20, 367, 1992
A:Title: Complete nucleotide sequence of a highly infectious avian leukosis virus.
A:Reference number: S35427; MUID:92158628
A:Accession: S35429
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-895 <BIE>
A:Cross-references: EMBL:M37980; NID:g210272; PIDN:AAA91269.1; PID:g210275
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1990
C:Genetics:
C:Superfamily: pol polyprotein
C:Keywords: polyprotein; reverse transcriptase

Query Match 41.2%; Score 395; DB 2; Length 895;
Best Local Similarity 46.4%; Pred. No. 8e-29;
Matches 83; Conservative 28; Mismatches 66; Indels 2; Gaps 2;

Qy 1 FTIPLAEOCEKFAFTIPAINNKEPATFQKVLPGQMLNSPTTCQTFVGRALQPVDRKF 60
```



```

Db 113 FSIPLAEQDCEKFAFTLPSPVNNQAPARRFQWKVLPQGMTCSPICQLVGVQVLEPLRLKH 172
QY 61 SDCVIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLATASDKIQSTSTPHYLGMQIEN 120
Db 173 PSRLMLHYMDLLAASHHDGLEAAGEEVISTLERAGFTISPDKVRQEPGVQVLYGLKGS 232
QY 121 RKIKPKQIEIRKDTLKTLDNFQKLLGDINWIRPTLGIPTVYAMNLSILRGSDSLNSKR 179
Db 233 TYVAPVGL-VAEPRIATLWDVQKLVGSLQWLRPALGIPPLRMGMFPFYEQLRG-SDPNEAR 289

RESULT 13
pol polyprotein - Rous sarcoma virus
N:Contains: endonuclease (EC 3.1.1.-); RNA-directed DNA polymerase (EC 2.7.7.49)
C:Species: Rous sarcoma virus
A:Note: host Gallus gallus (chicken)
C:Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 16-Jun-2000
C:Accession: A03955; S26418
R:Schwartz, D.; Tizard, R.; Gilbert, W.
Cell 32, 853-869, 1983
A:Title: Nucleotide sequence of Rous sarcoma virus.
A:Reference number: A90834; MUID:83155662
A:Accession: A03955
A:Molecule type: genomic RNA
A:Residues: 1-895 <SCH>
A:Cross-references: GB:V01197; NID:961695; PID:g1335587
A:Experimental source: strain Prague C
A:Note: as a result of base variations, a different version of this sequence may exist
R:Kashuba, V.I.; Rynditch, A.V.; Dostalova, V.; Hlozanek, I.; Zubak, S.V.; Kavsan, V.M.
submitted to the EMBL Data Library, September 1992
A:Description: Molecular cloning and DNA sequence analysis of duck-adapted variant of R
A:Reference number: S26417
A:Accession: S26418
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-47, 'V', 49-303, 'Q', 305-497, 'A', 499-565, 'K', 567-672, 'A', 674-895 <KAS>
A:Cross-references: EMBL:X68524; NID:g61903; PIDN:CAA48535.1; PID:g1334937
C:Comment: This protein is synthesized as a gag-pol polyprotein.
C:Genetics:
A:Gene: pol
C:Superfamily: pol polyprotein
C:Keywords: endonuclease; hydrolase; nucleotidyltransferase; polyprotein; reverse trans

Query Match 41.1%; Score 394; DB 1; Length 895;
Best Local Similarity 45.8%; Pred. No. 9.9e-29;
Matches 82; Conservative 29; Mismatches 66; Indels 2; Gaps 2;

QY 1 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGMNLSPTICQTFVGRALQPVDRKF 60
Db 113 FSIPLAEQDREAFATLPSPVNNQAPARRFQWKVLPQGMTCSPICQLVGVQVLEPLRLKH 172
QY 61 SDCVIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLATASDKIQSTSTPHYLGMQIEN 120
Db 173 PSRLMLHYMDLLAASHHDGLEAAGEEVISTLERAGFTISPDKVRQEPGVQVLYGLKGS 232
QY 121 RKIKPKQIEIRKDTLKTLDNFQKLLGDINWIRPTLGIPTVYAMNLSILRGSDSLNSKR 179
Db 233 TYVAPVGL-VAEPRIATLWDVQKLVGSLQWLRPALGIPPLRMGMFPFYEQLRG-SDPNEAR 289

RESULT 14
G48613
pol polyprotein - myeloblastosis-associated virus (strain MAV-1(N))
C:Species: myeloblastosis-associated virus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Feb-1998
C:Accession: G48613
R:Joliot, V.; Borouhgs, K.; Lasserre, F.; Crochet, J.; Dambrine, G.; Smith, R.E.; Perbal
Virology 195, 812-819, 1993
A:Title: Pathogenic potential of myeloblastosis-associated virus: implication of env pro
A:Reference number: A48613; MUID:93331743
```

```

A:Accession: G48613
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-896 <JOL>
A:Cross-references: GB:U10922
C:Superfamily: pol polyprotein

Query Match 39.7%; Score 381; DB 2; Length 896;
Best Local Similarity 45.3%; Pred. No. 1.6e-27;
Matches 81; Conservative 29; Mismatches 67; Indels 2; Gaps 2;

QY 1 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGMNLSPTICQTFVGRALQPVDRKF 60
Db 114 FSIPLAEQDREAFATLPSPVNNQAPARRFQWKVLPQGMTCSPICQLVGVQVLEPLRLKH 173
QY 61 SDCVIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLATASDKIQSTSTPHYLGMQIEN 120
Db 174 PSRLMLHYMDLLAASHHDGLEAAGEEVISTLERAGFTISPDKVRQEPGVQVLYGLKGS 233
QY 121 RKIKPKQIEIRKDTLKTLDNFQKLLGDINWIRPTLGIPTVYAMNLSILRGSDSLNSKR 179
Db 234 TYVAPVGL-VAEPRIATLWDVQKLVGSLQWLRPALGIPPLRMGMFPFYEQLRG-SDPNEAR 290

RESULT 15
E48613
pol polyprotein - myeloblastosis-associated virus (strain MAV-2(O)p9)
C:Species: myeloblastosis-associated virus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Feb-1998
C:Accession: E48613
R:Joliot, V.; Borouhgs, K.; Lasserre, F.; Crochet, J.; Dambrine, G.; Smith, R.E.; Per
Virology 195, 812-819, 1993
A:Title: Pathogenic potential of myeloblastosis-associated virus: implication of env
A:Reference number: A48613; MUID:93331743
A:Accession: E48613
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-896 <JOL>
A:Cross-references: GB:U10923
C:Superfamily: pol polyprotein

Query Match 39.7%; Score 381; DB 2; Length 896;
Best Local Similarity 45.3%; Pred. No. 1.6e-27;
Matches 81; Conservative 29; Mismatches 67; Indels 2; Gaps 2;

QY 1 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGMNLSPTICQTFVGRALQPVDRKF 60
Db 114 FSIPLAEQDREAFATLPSPVNNQAPARRFQWKVLPQGMTCSPICQLVGVQVLEPLRLKH 173
QY 61 SDCVIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLATASDKIQSTSTPHYLGMQIEN 120
Db 174 PSRLMLHYMDLLAASHHDGLEAAGEEVISTLERAGFTISPDKVRQEPGVQVLYGLKGS 233
QY 121 RKIKPKQIEIRKDTLKTLDNFQKLLGDINWIRPTLGIPTVYAMNLSILRGSDSLNSKR 179
Db 234 TYVAPVGL-VAEPRIATLWDVQKLVGSLQWLRPALGIPPLRMGMFPFYEQLRG-SDPNEAR 290

Search completed: April 9, 2002, 16:58:16
Job time: 250 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 16:57:29 ; Search time 34.73 Seconds
(without alignments)
117.927 Million cell updates/sec

Title: US-09-490-700-41
Perfect score: 959
Sequence: 1 FTPLAEDCKEFAFTIPAI.....SNLFSILRGDSLNKSRMLT 182

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	274.5	28.6	3080	6	5223423-4
2	270.5	28.2	1055	2	US-08-659-251-5
3	270.5	28.2	1055	4	US-09-256-490-5
4	270.5	28.2	1055	5	PCT-US96-11445-5
5	238.5	24.9	1150	4	US-09-238-303-9
6	237.5	24.8	314	1	US-08-589-446-6
7	237.5	24.8	314	1	US-08-444-882-6
8	237.5	24.8	314	2	US-08-389-459A-6
9	237.5	24.8	314	3	US-08-987-867A-6
10	237.5	24.8	562	4	US-09-117-217-14
11	237.5	24.8	913	2	US-07-743-357-22
12	237.5	24.8	1005	2	US-07-743-357-1
13	237.5	24.8	1016	2	US-07-743-357-4
14	237.5	24.8	1016	2	US-07-743-357-5
15	234.5	24.5	427	4	US-09-690-265-1
16	234.5	24.5	1003	2	US-07-743-357-9
17	234.5	24.5	1015	3	US-08-463-210-9
18	234.5	24.5	1015	4	US-09-124-900-3
19	234.5	24.5	1016	2	US-07-743-357-2
20	231.5	24.1	261	6	5320958-3
21	229.5	23.9	1004	2	US-07-743-357-7
22	228.5	23.8	913	2	US-07-743-357-6
23	227.5	23.7	1016	2	US-07-743-357-3
24	227	23.7	327	4	US-08-679-493A-71
25	226.5	23.6	260	6	5320958-4
26	222	23.1	995	5	PCT-US95-04910-14
27	219.5	22.9	1003	2	US-07-743-357-8

28	209.5	21.8	1003	2	US-07-743-357-10	Sequence 10, Appl
29	207.5	21.6	327	4	US-08-679-493A-72	Sequence 72, Appl
30	163	17.0	65	6	5320958-17	Patent No. 5320958
31	141.5	14.8	237	4	US-08-679-493A-73	Sequence 73, Appl
32	140.5	14.7	86	6	5320958-15	Patent No. 5320958
33	131.5	13.7	1203	4	US-09-075-272-4	Sequence 4, Appl
34	120.5	12.6	66	6	5320958-18	Patent No. 5320958
35	119.5	12.5	379	4	US-09-603-185-6	Sequence 6, Appl
36	119.5	12.5	1079	2	US-08-929-967-8	Sequence 8, Appl
37	114.5	11.9	665	2	US-08-929-967-7	Sequence 7, Appl
38	104.5	10.9	69	3	US-08-851-843A-16	Sequence 16, Appl
39	104.5	10.9	69	4	US-08-854-050-16	Sequence 16, Appl
40	97	10.1	1871	2	US-08-694-869-1	Sequence 1, Appl
41	97	10.1	1871	3	US-09-349-546-1	Sequence 1, Appl
42	95.5	10.0	65	6	5320958-16	Patent No. 5320958
43	89	9.3	67	6	5320958-14	Patent No. 5320958
44	78	8.1	410	2	US-08-741-327E-15	Sequence 15, Appl
45	76.5	8.0	845	1	US-08-416-950-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
5223423-4
; Patent No. 5223423
; APPLICANT: FRANCHINI, GENOVEFFA; WONG-STAAL, FLOSSIE;
; GALLO, ROBERT
; TITLE OF INVENTION: CHARACTERIZATION OF REPLICATION COMPETENT
; HUMAN IMMUNODEFICIENCY TYPE 2 PROVIRAL CLONE HIV-2 SBL/ISY
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/331,212
; FILING DATE: 03-31-1989
; SEQ ID NO:4:
LENGTH: 3080
5223423-4

Query Match 28.6%; Score 274.5; DB 6; Length 3080;
Best Local Similarity 37.7%; Pred. No. 2.9e-24;
Matches 66; Conservative 31; Mismatches 69; Indels 9; Gaps 4;

Qy	1	FTPLAEDCKEFAFTIPAINNKEPATRFQWKLVPQGLNSPTICQTFVGRALQPVDRKF	60
Db	882	FSIPLYEDFRQYTAFTLPSVNAEPGRKYIVKLPQGWKSPAIQYTMROVLEPFRKAN	941
Qy	61	SDCVIIHFDDILCAAEKDKLIDCYFLPAEVANA-GLATASOKIQSTTFHYLGMQIE	119
Db	942	PDVIVQYMDLILASDRTDLEHDKVVLQKLKELLNGLGFSTPDEKFKDPPYQWNGYELW	1001
Qy	120	NRKIKPKQIEIRKDTLTLNDFQKLGDINW---IRPTLGIPTVAMSLFSLRG	171
Db	1002	PTKWKLQIKLPQKEVWTVNDIQLGVINWAAQIYP--GIKT---KHLCKLIRG	1051

RESULT 2
US-08-659-251-5
; Sequence 5, Application US/08659251
; Patent No. 5883081
; GENERAL INFORMATION:
; APPLICANT: Kraus, Guenter
; APPLICANT: Wong-Staal, Flossie
; APPLICANT: Talbott, Randy
; APPLICANT: Poeschia, Eric
; TITLE OF INVENTION: Isolation of No. 5883081el HIV-2 Proviruses
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA

```

4
PCRT-US96-111445-5
; Sequence 5, Application PC/TUS96111445
; GENERAL INFORMATION:
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Isolation of Novel HIV-2 Proviruses
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Robbins, Berliner & Carson
; STREET: 201 N. Figueroa Street, 5th Floor
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90012-2628
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
;

```

```

QY 1 FTPLAEOCEKFAFTTIPAINNKEPATRFQWKVLPOQMLNSPTICQTFVGRALQPVDRKF 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 296 FTPLDPDYAPYTAFTLTPKINNPGSGRFFVCMGLPOQGWLSPLIYOSTLNNILKPFREQH 355
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 SDCVIIHYFDIILCAAEKTKDLIDCYTFLPAEVANA---GLAIADSKIQTSTPEHYLGM 116
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 356 PEIDLVOYMDDIVIGSDLGHKE---HKQIVVEERKLLLLWGFETPEDKLGEOQPYKMGY 412
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 117 QIENNRKTKPKQIEIRKDTLTKLNDFOKLGDINWIRPTLGIPTYAMNLSFIILRGSDLN 176
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 413 ELYPRKWTQTKELIPEPTNELQRLVGIINWSSQI--IPGLRIKALTNMKNQALD 470
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 177 SKRWLT 182
      |||||
Db 471 SKRWT 476
      |||||

RESULT 6
US-08-589-446-6
; Sequence 6, Application US/08589446
; Patent No. 5614413
; GENERAL INFORMATION:
; APPLICANT: MORROW, Casey D.
; TITLE OF INVENTION: ENCAPSIDATED POLIOVIRUS NUCLEIC
; TITLE OF INVENTION: ACID AND METHODS OF MAKING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/589,446
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,009
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Geary III, William C.
; REGISTRATION NUMBER: 31,359
; REFERENCE/DOCKET NUMBER: UAG-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 314 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-589-446-6

Query Match 24.8%; Score 237.5; DB 1; Length 314;
Best Local Similarity 33.0%; Pred. No. 3.7e-21;
Matches 62; Conservative 32; Mismatches 81; Indels 13; Gaps

QY 1 FTPLAEOCEKFAFTTIPAINNKEPATRFQWKVLPOQMLNSPTICQTFVGRALQPVDRK 59
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94 FSVPL-DEDRPKYTAFTTIPINNTEPGIRQYNVLPQWKGSPAIFOSSTKILEPFRKQ 157
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 60 FSDCVIIHYFDIILCAAE-----KKDLIDCYTFLPAEVANAGLAIASDRKIQTSTPEHYL 114
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 153 NPDIVYQYMDLLVGSDELTGQHRKTIEE-----LRQHLLRWGLTTPDKKHQKEPPFLWM 208
Qy 115 GMOIENRKIPQKIEIRKDTLTKLNDQKLLGDNWIRPTLGIPTYAMSNLFSILRGDS 174
Db 209 GYELHPDKWTVPVLPKDSWTVNDIQKLVGLKNWASQI--YPGIKVROLCKLLRGTKA 266
Qy 175 LNSKRLMT 182
Db 267 LTEVIPLT 274

RESULT 7
US-08-444-882-6
; Sequence 6, Application US/08444882
; Patent No. 5622705
; GENERAL INFORMATION:
; APPLICANT: Morrow, Casey D.
; TITLE OF INVENTION: ENCAPSIDATED POLIOVIRUS NUCLEIC
; TITLE OF INVENTION: ACID AND METHODS OF MAKING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,882
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/087,009
; FILING DATE: 01-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Geary III, William C.
; REGISTRATION NUMBER: 31,359
; REFERENCE/DOCKET NUMBER: UAG-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 314 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-882-6

Query Match 24.8%; Score 237.5; DB 1; Length 314;
Best Local Similarity 33.0%; Pred. No. 3.7e-21;
Matches 62; Conservative 32; Mismatches 81; Indels 13; Gaps 5;
Qy 1 FTIPLAEDCEKF-AFTIPAINNKEPATRFQWKVLPQGMNLSPICQTFVGRALQPVDRK 59
Db 94 FSVPL-DEDFRKYTAFTIP SINNETPGIRYQYNNVLPQGWKSPAIQSSMTKILEPFRKQ 152
Qy 60 FSDCYIIHYFDILCAAEET-----KDKLIDCYTFLPAEVANAGLAIASDKIOTSTPPEHYL 114
Db 153 NPDIVYQYMDLLVGSDELTGQHRKTIEE-----LRQHLLRWGLTTPDKKHQKEPPFLWM 208
Qy 115 GMOIENRKIPQKIEIRKDTLTKLNDQKLLGDNWIRPTLGIPTYAMSNLFSILRGDS 174
Db 209 GYELHPDKWTVPVLPKDSWTVNDIQKLVGLKNWASQI--YPGIKVROLCKLLRGTKA 266
Qy 175 LNSKRLMT 182

Db 267 LTEVIPLT 274
RESULT 8
US-08-389-459A-6
; Sequence 6, Application US/08389459A
; Patent No. 5817512
; GENERAL INFORMATION:
; APPLICANT: Morrow, Casey D. and Porter, Donna, C.
; TITLE OF INVENTION: ENCAPSIDATED RECOMBINANT POLIOVIRUS
; TITLE OF INVENTION: NUCLEIC ACID AND METHODS OF MAKING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,459A
; FILING DATE: 15-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/087,009
; FILING DATE: 01-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Silveri, Jean M.
; REGISTRATION NUMBER: 39,030
; REFERENCE/DOCKET NUMBER: UAG-004CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 314 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-389-459A-6

Query Match 24.8%; Score 237.5; DB 2; Length 314;
Best Local Similarity 33.0%; Pred. No. 3.7e-21;
Matches 62; Conservative 32; Mismatches 81; Indels 13; Gaps 5;
Qy 1 FTIPLAEDCEKF-AFTIPAINNKEPATRFQWKVLPQGMNLSPICQTFVGRALQPVDRK 59
Db 94 FSVPL-DEDFRKYTAFTIP SINNETPGIRYQYNNVLPQGWKSPAIQSSMTKILEPFRKQ 152
Qy 60 FSDCYIIHYFDILCAAEET-----KDKLIDCYTFLPAEVANAGLAIASDKIOTSTPPEHYL 114
Db 153 NPDIVYQYMDLLVGSDELTGQHRKTIEE-----LRQHLLRWGLTTPDKKHQKEPPFLWM 208
Qy 115 GMOIENRKIPQKIEIRKDTLTKLNDQKLLGDNWIRPTLGIPTYAMSNLFSILRGDS 174
Db 209 GYELHPDKWTVPVLPKDSWTVNDIQKLVGLKNWASQI--YPGIKVROLCKLLRGTKA 266
Qy 175 LNSKRLMT 182
Db 267 LTEVIPLT 274

RESULT 9
US-08-987-867A-6

; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: HXB2
US-07-743-357-22

Query Match 24.8%; Score 237.5; DB 2; Length 913;
Best Local Similarity 33.0%; Pred. No. 1.8e-20;
Matches 62; Conservative 32; Mismatches 81; Indels 13; Gaps 5;

Qy 1 FTIPLAEDCEKFAFTIPAINNKEPATRFQWKVLPQGMNLSPICQTFVGRALQPVDRK 59
Db 180 FSVPL-DEDFRKYTAFTIPSNINNETPGIRYQYNVLPQGWKSPAIQSSMTKILEPFRKQ 238
Qy 60 FSDCYIIHYFDDILCAET-----KDKLIDCYTFLPAEVANAGLAIASDKIQSTPFFHYL 114
Db 239 NPDIVIYQYMDLVVGSDEIGQHRKTEE---LRQHLLRWGLTTPDKKHQKEPFFLWM 294
Qy 115 GMQENRKIKPKQKEIRKDTLTKLNDFOKLLGDNINWIRPTLIGIPTYAMNSLFSILRGDSD 174
Db 295 GYELHPDKWTQPIVLPEKDSMTVNDIQKLVGKLNWASQI--YPGIKVRQLCKLLRGTKA 352
Qy 175 LNSKRMLT 182
Db 353 LTVIPLT 360

RESULT 12
US-07-743-357-1
; Sequence 1, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REFERENCE/DOCKET NUMBER: 28,584
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1005 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein

; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: HXB2
US-07-743-357-1

Query Match 24.8%; Score 237.5; DB 2; Length 1005;
Best Local Similarity 33.0%; Pred. No. 2.le-20;
Matches 62; Conservative 32; Mismatches 81; Indels 13; Gaps 5;

Qy 1 FTIPLAEDCEKFAFTIPAINNKEPATRFQWKVLPQGMNLSPICQTFVGRALQPVDRK 59
Db 272 FSVPL-DEDFRKYTAFTIPSNINNETPGIRYQYNVLPQGWKSPAIQSSMTKILEPFRKQ 330
Qy 60 FSDCYIIHYFDDILCAET-----KDKLIDCYTFLPAEVANAGLAIASDKIQSTPFFHYL 114
Db 331 NPDIVIYQYMDLVVGSDEIGQHRKTEE---LRQHLLRWGLTTPDKKHQKEPFFLWM 386
Qy 115 GMQENRKIKPKQKEIRKDTLTKLNDFOKLLGDNINWIRPTLIGIPTYAMNSLFSILRGDSD 174
Db 387 GYELHPDKWTQPIVLPEKDSMTVNDIQKLVGKLNWASQI--YPGIKVRQLCKLLRGTKA 444
Qy 175 LNSKRMLT 182
Db 445 LTVIPLT 452

RESULT 13
US-07-743-357-4
; Sequence 4, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REFERENCE/DOCKET NUMBER: 28,584
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1016 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein


```

; HYPOTHETICAL: NO
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: PV22
; US-07-743-357-4

```

```

Query Match 24.8%; Score 237.5; DB 2; Length 1016;
Best Local Similarity 33.0%; Pred. No. 2.1e-20;
Matches 62; Conservative 32; Mismatches 81; Indels 13; Gaps 5;

```

```

QY 1 FTPLAQDCEKF-AFTIPAINNKEPATRFQWKVLPQGMNLSPTICOTFVGRALQPVDRK 59
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 283 FSVPL-DEDFRKYTAFTIPAINNKEPATRFQWKVLPQGMNLSPTICOTFVGRALQPVDRK 341
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
QY 60 FSDCYIIHYFDDILCAAE-----KDKLIDCYTFLPAEVANAGLAIASDKIQTSTPPEHYL 114
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 342 NPDIIVYQYMDLLVSGDLEIGQHRKIEE-----LRQHLRWGLTTPDKKHQKEPPFLWM 397
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
QY 115 GMOJENRKIKPQKIEIRKDTLKTLDNFQKLGIDINWIRPTLIGIPTYAMSNLFSILRGDSD 174
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 398 GYELHPDKWTVPVLPPEKDSWTVDIQKLVGKLNWASQI--YPIKVRQLCKLLRGTKA 455
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
QY 175 LNSKRMLT 182
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 456 LVEIPLT 463

```

```

RESULT 14
US-07-743-357-5
; Sequence 5, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1016 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: Internal

```

```

; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: BRU
; US-07-743-357-5

```

```

Query Match 24.8%; Score 237.5; DB 2; Length 1016;
Best Local Similarity 33.0%; Pred. No. 2.1e-20;
Matches 62; Conservative 32; Mismatches 81; Indels 13; Gaps 5;

```

```

QY 1 FTPLAQDCEKF-AFTIPAINNKEPATRFQWKVLPQGMNLSPTICOTFVGRALQPVDRK 59
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 283 FSVPL-DEDFRKYTAFTIPAINNKEPATRFQWKVLPQGMNLSPTICOTFVGRALQPVDRK 341
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
QY 60 FSDCYIIHYFDDILCAAE-----KDKLIDCYTFLPAEVANAGLAIASDKIQTSTPPEHYL 114
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 342 NPDIIVYQYMDLLVSGDLEIGQHRKIEE-----LRQHLRWGLTTPDKKHQKEPPFLWM 397
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
QY 115 GMOJENRKIKPQKIEIRKDTLKTLDNFQKLGIDINWIRPTLIGIPTYAMSNLFSILRGDSD 174
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 398 GYELHPDKWTVPVLPPEKDSWTVDIQKLVGKLNWASQI--YPIKVRQLCKLLRGTKA 455
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
QY 175 LNSKRMLT 182
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 456 LVEIPLT 463

```

```

RESULT 15
US-09-690-265-1
; Sequence 1, Application US/09690265
; Patent No. 6300351
; GENERAL INFORMATION:
; APPLICANT: Uckun, Fatih A.
; APPLICANT: Mao, Chen
; TITLE OF INVENTION: BETA-FLUOROETHYL THIOUREA COMPOUNDS AND USE
; FILE REFERENCE: 12152.8USCA
; CURRENT APPLICATION NUMBER: US/09/690,265
; CURRENT FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: US 09/205,167
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; US-09-690-265-1

```

```

Query Match 24.5%; Score 234.5; DB 4; Length 427;
Best Local Similarity 32.4%; Pred. No. 1.4e-20;
Matches 61; Conservative 33; Mismatches 81; Indels 13; Gaps 5;

```

```

QY 1 FTPLAQDCEKF-AFTIPAINNKEPATRFQWKVLPQGMNLSPTICOTFVGRALQPVDRK 59
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 116 FSVPL-DEDFRKYTAFTIPAINNKEPATRFQWKVLPQGMNLSPTICOTFVGRALQPVDRK 174
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
QY 60 FSDCYIIHYFDDILCAAE-----KDKLIDCYTFLPAEVANAGLAIASDKIQTSTPPEHYL 114
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 175 NPDIIVYQYMDLLVSGDLEIGQHRKIEE-----LRQHLRWGLTTPDKKHQKEPPFLWM 230
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
QY 115 GMOJENRKIKPQKIEIRKDTLKTLDNFQKLGIDINWIRPTLIGIPTYAMSNLFSILRGDSD 174
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 231 GYELHPDKWTVPVLPPEKDSWTVDIQKLVGKLNWASQI--YPIKVRQLCKLLRGTKA 288
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
QY 175 LNSKRMLT 182
   ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 289 LVEIPLT 296

```

Search completed: April 9, 2002, 16:57:30
Job time: 249 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 16:56:45 ; Search time 72.35 Seconds
(without alignments)
186.335 Million cell updates/sec

Title: US-09-490-700-41

Perfect score: 959

Sequence: 1 FTIPLAEQCEKFAFTIPAI.....SNLFSILRGSDLSKRMILT 182

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A_Geneseq_1101.*
- 1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.*
 - 22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred.. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	959	100.0	182	20	AAW95695 Human endogenous r
2	959	100.0	182	20	AAW97748 Human endogenous r
3	525	54.7	870	14	AAK31348 Jaagsiekte retrovi
4	486.5	50.7	197	22	AAI18128 Peptide #4562 enco
5	486.5	50.7	197	22	AAK30633 Peptide #4670 enco
6	486.5	50.7	197	22	AAK05759 Peptide #4441 enco
7	464	48.4	1755	20	AAV41139 Mouse mammary tumo
8	461	48.1	734	20	AAV52087 Human retrovirus-5
9	405	42.2	775	6	AAK50121 Sequence of a poly
10	405	42.2	775	8	AAK70417 Polypeptide with I
11	397	41.4	578	21	AAK12991 MAV reverse transc

12	397	41.4	832	21	AAK12992 MAV reverse transc
13	397	41.4	895	21	AAK12989 Full length avian
14	397	41.4	896	21	AAK12990 MAV reverse transc
15	387	40.4	895	16	AAK80522 Rous associated vi
16	284	29.6	603	22	AAK16982 Peptide #3416 enco
17	284	29.6	603	22	AAK29472 Peptide #3509 enco
18	284	29.6	603	22	AAK04685 Peptide #3367 enco
19	281.5	29.4	1014	9	AAK08010 Sequence of pol pr
20	281.5	29.4	1027	9	AAK81773 Sequence encoded b
21	281.5	29.4	1036	13	AAK20599 ROD HIV-2 polymera
22	281.5	29.4	1036	21	AAK51978 HIV-2 ROD isolate
23	279.5	29.1	1060	13	AAK23366 SiVmac239 pol gene
24	279.5	29.1	1060	20	AAK89314 SiVmac239 genome p
25	276.5	28.8	1055	21	AAK12993 HIV-2 reverse tran
26	274.5	28.6	1056	9	AAK81783 Sequence encoded b
27	274.5	28.6	3080	10	AAK933285 Sequence of clone
28	274.5	28.6	3210	9	AAK81771 Deduced sequence e
29	271.5	28.3	1056	9	AAK80809 Sequence of pol pr
30	270.5	28.2	1055	18	AAK13055 HIV-2 provirus-enc
31	263.5	27.5	1143	21	AAK35755 EIAV pol gene prod
32	261	27.2	196	22	AAK51187 PERV-MSN1 BTP-4 re
33	260.5	27.2	1035	11	AAK04025 Pol gene product o
34	238.5	24.9	372	22	AAK81068 HIV protease and r
35	238	24.8	82	22	AAK35533 Peptide #9570 enco
36	237.5	24.8	917	16	AAK64974 HIV-1 pol protein.
37	237.5	24.8	313	17	AAK00180 Human immunodefici
38	237.5	24.8	850	22	AAK04791 Human tPA leader s
39	237.5	24.8	875	22	AAK04793 HIV-1 reverse tran
40	237.5	24.8	902	21	AAK12995 AcNPV-HIVK-pol pr
41	237.5	24.8	912	11	AAK08053 Sequence of LAV vi
42	237.5	24.8	1003	7	AAK60420 Sequence encoded b
43	237.5	24.8	1003	8	AAK70861 Sequence of revers
44	237.5	24.8	1012	7	AAK61507 HTLV-III pol prote
45	237.5	24.8	1012	20	AAK90176

ALIGNMENTS

RESULT 1
AAW95695
ID AAW95695 standard; Protein; 182 AA.
XX AC AAW95695;
XX AC
XX DT 08-JUN-1999 (first entry)
XX DT
XX DE Human endogenous retrovirus IDMK1.2-22 pol protein.
XX DE
XX KW IDDM; insulin-dependent diabetes mellitus; endogenous retrovirus;
XX KW SAQ; superantigen; provirus; autoimmune disease; type 1 diabetes;
XX KW diagnosis; ss.
XX OS Homo sapiens.
XX PN EP893691-A1.
XX PD 27-JAN-1999.
XX PF 23-JUL-1997; 97EP-0401773.
XX PR 23-JUL-1997; 97EP-0401773.
XX PA (MACH/) MACH B F.
XX PI Conrad B, Mach B;
XX DR WPI; 1999-097928/09.
XX PT Diagnosing human autoimmune disease by detecting retrovirus with
XX PT superantigen activity - new retrovirus associated with type 1
XX PT diabetes, its proviral DNA, and related vectors, transformed cells,
XX PT proteins, antibodies and specific binding agents, used for treating

PD	20-NOV-1992.	
XX		
PF	17-MAY-1991;	91FR-0006060.
XX		
PR	17-MAY-1991;	91FR-0006060.
XX		
PA	(INRM) INSERM INST NAT SANTE & RECH MED.	
XX		
PI	Querat GF, Verwoerd D, Vigne R, York D;	
XX		
WD	WPI; 1993-020250/03.	
N-PSDB:	AAQ35153.	
XX		
PT	New Jaagsiekte Retrovirus and characteristic nucleic acid - also	
PT	derived proteins, probes and antibodies, useful for in vitro	
PT	diagnosis and in vaccines	
XX		
PS	Claim 25; Page 39-42; 75pp; French.	
XX		
CC	JSRV causes epithelial carcinoma in ovine animals, partic. pulmonary	
CC	adenomatosa in sheep. The complete cDNA sequence of the JSRV genome	
CC	was prepared from an approx. 8.7kb band of poly-A RNA isolated from	
CC	semi-purified lung-lavage samples from infected sheep. The	
CC	invention includes the Pol amino acid sequence or any part of it	
CC	which is capable of specific immunological reaction with antibodies	
CC	directed against JSRV.	
CC	See also AAR31346-7, AAR31349 and AAQ35153-Q35155.	
XX		
SQ	Sequence	870 AA;

Query Match	54.7%	Score 525;	DB 14;	Length 870;
Best Local Similarity	53.3%;	Pred. No. 2e-50;		
Matches 97;	Conservative 31;	Mismatches 54;	Indels 0;	Gaps 0;

QY	1	FTFLAEQDCEKFAFTIPAINNKSPATRFQOKVLPOGLMNSPTTCOTFVGALQPVDRKE	60
		: : : : : : : : : : : : : : :	
Db	126	ytiplapqdcckrfafspvnmfkepmqzyqrwlvpqgntnsptticqkfvaiaapvqrfr	185
QY	61	SDCVIIHYFDDILCAAETKDKLIDCYFTFLPAEVANAGIAISDXIKQTSTPHYLGQMGIN	120
		: : : : : : : : : : : : : :	
Db	186	pqlyvlvhyrmdilliahtdehilygafsilqkdhlsinglviaadekigtbfpynylgfslvp	245
QY	121	RKIKPQKIEKTRDKLTKNDFOKLLGDINWIRTPTLGIPTYAMSNLFSTLRGDSDLNSKRM	180
		: : : : : : : : : : : : : :	
Db	246	rvyntqlvkiqtqdhkhtlndiqkllgdnwlrpylkiptylqlpfldikgdsdpasprt	305
QY	181	LT	182
		: :	
Db	306	ls	307

RESULT	4	
AAM18128		
ID	AAM18128	standard; Protein; 197 AA.
XX	AC	
XX	AC	
XX	XX	
XX	XX	
DT	12-OCT-2001	(first entry)
XX		
DE	Peptide #4562	encoded by probe for measuring cervical gene expression.
XX		
KW	Probe; human; microarray;	gene expression; cervical epithelial cell;
KW	cervical cancer.	
XX		
OS	Homo sapiens.	
XX		
XX		
PN	WO200157278-A2.	
XX		
PD	09-AUG-2001.	
XX		
PF	30-JAN-2001;	2001WO-US00670.
XX		

04-FEB-2000; 2000US-0180312. PR
26-MAY-2000; 2000US-0207456. PR
30-JUN-2000; 2000US-0608408. PR
03-AUG-2000; 2000US-0632366. PR
21-SEP-2000; 2000US-0234687. PR
27-SEP-2000; 2000US-0236359. PR
04-OCT-2000; 2000GB-0024263. PR
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC. XX
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR; XX
PI XX
XX WPI; 2001-488901/53. DR
XX
XX Human genome-derived single exon nucleic acid probes useful for PT
PT analyzing gene expression in human cervical epithelial cells - XX
XX
XX Claim 27; SEQ ID No 22954; 487pp; English. PS
XX
XX The present invention relates to human single exon nucleic acid probes CC
CC (SENP; see AA110068-AA128459). The present sequence is a peptide encoded CC
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs CC
CC can be used to produce a single exon microarray, which can be used for CC
CC measuring human gene expression in a sample derived from human cervical CC
CC epithelial cells. By measuring gene expression, the probes are therefore CC
CC useful in grading and/or staging of diseases of the cervix, notably CC
CC cervical cancer. CC
CC Note: The sequence data for this patent did not form part of the printed CC
CC specification, but was obtained in electronic format directly from WIPO CC
CC at ftp.wipo.int/pub/published_pct_sequences. CC
XX
XX Sequence 197 AA; SQ

Query Match	50.7%;	Score	486.5;	DB	22;	Length	197;
Best Local Similarity	89.6%;	Pred.	No. 5.8e-47;				
Matches	95;	Conservative	2;	Mismatches	4;	Indels	5;
						Gaps	1;
Qy	1	FTPLAEDQCEKFAFTIPAINNKEPATRFQKVLPGQMLNSPTICQTFVGRALQPVROKF	50				
Db	97	ftplaeqdyekfaftipainnkepatrfqkwlpqgmlnspctlcq-----alqpvrdf	151				
Qy	61	SDCYIIHYFDDILCAAETKDKLIDCYTFELPAEVA NAGLAIASDKIQ	106				
Db	152	sdcyilhyvddilcaeamrdrklincytfelqaevanaglaia sddkik	197				
RESULT	5						
AAM30633							
ID	AAM30633	standard; Protein; 197	AA.				
XX							
AC	AAM30633;						
XX							
DT	17-OCT-2001	(first entry)					
XX							
DE	Peptide #4670	encoded by probe for measuring placental gene expression.					
XX							
KW	Probe; microarray; human; placenta; antenatal diagnosis;						
KW	genetic disorder.						
XX							
OS	Homo sapiens.						
XX							
PN	W0200157272-A2.						
XX							
PD	09-AUG-2001.						
XX							
XX	30-JAN-2001; 2001WO-US00663.						
XX							
PR	04-FEB-2000; 2000US-0180312.						
PR	26-MAY-2000; 2000US-0207456.						
PR	30-JUN-2000; 2000US-0608408.						
PR	03-AUG-2000; 2000US-0632366.						
PR	21-SEP-2000; 2000US-0234687.						

```

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 27; SEQ ID NO 30902; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP:
XX see AAI31315-AA157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders.
XX
XX Sequence 197 AA;

Query Match 50.7%; Score 486.5; DB 22; Length 197;
Best Local Similarity 89.6%; Pred. No. 5.8e-47;
Matches 95; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

.QY 1 FTPLAEQDCCKFAFTIPAINNKEPATRFQWKVLPQGMINSPTICQTFVGRALQPVDRKF 60
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
97 ftplaeqdyekftaipainnkepatrfqwkvlpggmnspticq-----alqpvrdfk 151

.QY 61 SDCVIIHHYFDDILCAAEKTKKLDICYTFLPAEVANAGLAIASDKIQ 106
Db ||||||| ||||||| :||||: ||||| ||||||| |||||||
152 sdcylhyvddilcaaeMrdklncylflqaevanaglaiaasdkik 197

RESULT 6
AA05759
ID AAM05759 standard; Protein; 197 AA.
XX
XX AAM05759;
XX
XX 09-OCT-2001 (first entry)
XX
XX Peptide #441 encoded by probe for measuring breast gene expression.
XX
XX Probe: human; breast disease; breast cancer; development disorder;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX WO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression

PT in a human breast -
XX
XX Claim 27; SEQ ID NO 14499; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes
XX (see AAI00010-AA110067). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for measuring human gene expression in
XX a human breast sample, where the probe hybridises at high stringency to a
XX nucleic acid expressed in the human breast. The probes are useful for
XX predicting, diagnosing, grading, staging, monitoring and prognosing
XX diseases of the human breast, particularly those diseases with polygenic
XX aetiology. The diseases include: breast cancer, disorders of development,
XX inflammatory diseases of the breast, fibrocystic changes, proliferative
XX breast disease and non-carcinoma tumours.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 197 AA;

Query Match 50.7%; Score 486.5; DB 22; Length 197;
Best Local Similarity 89.6%; Pred. No. 5.8e-47;
Matches 95; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

.QY 1 FTPLAEQDCCKFAFTIPAINNKEPATRFQWKVLPQGMINSPTICQTFVGRALQPVDRKF 60
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
97 ftplaeqdyekftaipainnkepatrfqwkvlpggmnspticq-----alqpvrdfk 151

.QY 61 SDCVIIHHYFDDILCAAEKTKKLDICYTFLPAEVANAGLAIASDKIQ 106
Db ||||||| ||||||| :||||: ||||| ||||||| |||||||
152 sdcylhyvddilcaaeMrdklncylflqaevanaglaiaasdkik 197

RESULT 7
AA41139
ID AAY41139 standard; Protein; 1755 AA.
XX
XX AAY41139;
XX
XX 24-JAN-2000 (first entry)
XX
XX Mouse mammary tumor virus (MMTV) gag-pol-pro polypeptide.
XX
XX Immune response; mouse mammary tumor virus; MMTV; immunomodulatory;
XX epitope; passive immunotherapy; gag; pol; pro.
XX
XX Mouse mammary tumor virus.
XX
XX OS
XX WO951268-A1.
XX
XX PN
XX 14-OCT-1999.
XX
XX PD
XX 08-APR-1999; 99WO-US07712.
XX
XX PF
XX 08-APR-1998; 98US-0057003.
XX
XX PR
XX (ITBI-) INT BIOIMMUNE SYSTEMS INC.
XX
XX PA
XX Stewart THM, Gershoni J;
XX
XX PI
XX WPI; 1999-620171/53.
XX
XX DR
XX N-PSDB; AAZ23193.
XX
XX DR
XX Novel immunomodulatory compositions of mouse mammary tumor virus (MMTV)
XX antigens used for treating or preventing MMTV infections -
XX
XX PT
XX Disclosure; Fig 2; 92pp; English.
XX
XX PS
XX
XX The invention provides a method for inducing an immune response to mouse
XX mammary tumor virus (MMTV) in a human subject. The method comprises
XX administering to the subject an immunomodulatory composition comprising a
XX pharmaceutical carrier and at least one MMTV antigen (or a discontinuous

```

CC epitope of MMTV) in an amount sufficient to elicit an increase in a MMTV
 CC specific cellular or humoral response. The compositions and methods of
 CC the invention are useful for treating or preventing a human disease or
 CC disorder caused by MMTV, or a biological or chemical agent that reacts
 CC with an antibody directed towards an MMTV antigen. The compositions and
 CC antibodies may be used in passive immunotherapy. The immunomodulatory
 CC compositions can be used in a preventative manner for those subjects not
 CC exposed to MMTV or MMTV antigen reactive agents. The MMTV antigens also
 CC have use in diagnostic assays. The present sequence represents the
 CC amino acid sequence of the MMTV gag-pol-pro polypeptide.

XX Sequence 1755 AA;

Query Match 48.4%; Score 464; DB 20; Length 1755;
 Best Local Similarity 50.0%; Pred. No. 4.4e-43;
 Matches 91; Conservative 30; Mismatches 61; Indels 0; Gaps 0;

QY 1 FTPLAQDECKEFAFTIPAINNKEPATRFQWKVLPQGLMNSPTTCQTFVGRALQPVDRDKF 60
 Db 976 fnklhpdcckrfafsvpsnfkrpyqrfqkvlpqgmknspcticqkfdvkailltrdky 1035
 QY 61 SDCYIIHYFDDILCAAETKDKLIDCYTFLEPAEVANAGLAIASDKIQSTPFFHYLGMQIEN 120
 Db 1036 qdsyivhymddillahpsrsivdeiltsmiqalnkhglvstekiqlkqydnlkylgthlqg 1095
 QY 121 RKIKPQKIEIRKDKTLNDFQKLLGDNINWIRPTLGIPTVAMSNLFSILRGDSDLNSKRM 180
 Db 1096 dsvsqklqtrtdklrlnndfqlglninwlrpfklittgeklpfeilngdpsnistrk 1155
 QY 181 LT 182
 Db 1156 lt 1157

RESULT 8

AAY52087
 ID AAY52087 standard; Protein; 734 AA.

AC AAY52087;

XX 24-JAN-2000 (first entry)

DE Human retrovirus-5 pol amino acid sequence.

XX HRV-5; Human retrovirus-5; gag; pro; pol; nucleoprotein; polymerase;
 KW recombination; PCR primer; detect; therapy; antibody; vaccine; diagnosis;
 KW prognosis; rheumatoid arthritis; osteoarthritis; Sjogren's disease;
 KW systemic lupus erythematosus; inflammatory bowel disease;
 KW autoimmune disease; inflammatory disease.

XX Human retrovirus-5.

XX WO9950285-A2.

XX 07-OCT-1999.

XX 26-MAR-1999; 99WO-GB000956.

XX 27-MAR-1998; 98GB-0006649.

PR 08-JAN-1999; 99GB-0000409.

XX (CANC-) CANCER RES INST.

PA (KENN-) KENNEDY INST RHEUMATOLOGY MATHILDA & TER.

XX Griffiths DJ, Weiss RA, Venables PJW, Boyd MT;

XX WPI; 1999-601321/51.

DR N-PSDB; AA232523.

XX Human retrovirus-5, its nucleic acid and derived proteins, useful for
 PT the treatment, diagnosis and prevention of autoimmune and inflammatory
 PT diseases -

XX Claim 10; Fig 12; 105pp; English.

XX This is the human retrovirus-5 HRV-5 pol amino acid sequence. The full
 CC length HRV-5 nucleotide sequence AA232523, encodes the gag, pol and pro
 CC genes of HRV-5. The gag gene encodes the components of the nucleoprotein
 CC of the virus. The pol gene codes for proteins involved in nucleic acid
 CC synthesis and recombination, and the pro gene gives rise to the
 CC protease protein. The HRV-5 nucleotide sequences are used in the
 CC invention to create PCR primers which can be used to detect HRV-5 in
 CC samples from patients. HRV-5 proviral DNA has been detected in
 CC inflamed joints, but not in normal synovium. HRV-5 nucleic acids may
 CC also be used to screen for specific inhibitors (potential therapeutic
 CC agents) and to produce recombinant polypeptides. The virus itself, when
 CC disabled, can be used as a gene therapy vector. HRV-5 polypeptides are
 CC used to raise antibodies (which may be used to detect the virus or as
 CC therapeutic inhibitor), to screen for modulators and in vaccines.
 CC Fragments of the HRV-5 nucleotide may be used as probes or primers for
 CC viral detection (for diagnosis or prognosis) and as sources of
 CC therapeutic antisense sequences. The various therapeutic agents can be
 CC used to treat rheumatoid arthritis, osteoarthritis, systemic lupus
 CC erythematosus, inflammatory bowel disease, Sjogren's syndrome and other
 CC inflammatory or autoimmune conditions.

XX Sequence 734 AA;

Query Match 48.1%; Score 461; DB 20; Length 734;
 Best Local Similarity 47.3%; Pred. No. 2.8e-43;
 Matches 86; Conservative 37; Mismatches 59; Indels 0; Gaps 0;

QY 1 FTPLAQDECKEFAFTIPAINNKEPATRFQWKVLPQGLMNSPTTCQTFVGRALQPVDRDKF 60
 Db 120 fsiplheqdiqrfaftvpsinhqgdkryekwlpqgmtnspalcqlvydqvavevrqgc 179
 QY 61 SDCYIIHYFDDILCAAETKDKLIDCYTFLEPAEVANAGLAIASDKIQSTPFFHYLGMQIEN 120
 Db 180 pkvgilhydmddilltaeseshlmeaykilllylekvgilqvapekikggevvyglkvts 239
 QY 121 RKIKPQKIEIRKDKTLNDFQKLLGDNINWIRPTLGIPTVAMSNLFSILRGDSDLNSKRM 180
 Db 240 ekvtplefeialdglqtlndfqlglninwlrpycklittedmmpsfnllegdaqidspr 299
 QY 181 LT 182
 Db 300 lt 301

RESULT 9

AAP50121
 ID AAP50121 standard; Protein; 775 AA.

AC AAP50121;

XX 27-SEP-1991 (first entry)

XX Sequence of a polypeptide exhibiting mammalian immunoglobulin
 DE binding factor activity (IBF) encoded by cDNA clone 23B6p10.2.

XX Immunoglobulin E-mediated disease; therapy; B-cell differentiation;
 KW immunoglobulin E-binding factor.

XX EPI55192-A.

XX 18-SEP-1985.

XX 15-MAR-1985; 85EP-0301834.

XX 16-MAR-1984; 84US-0590430.

XX (SCHE-) SCHERING BIOTECH CO.

PA (UYJO) JOHNS HOPKINS UNIV.

PA (SCHE) SCHERING-BIOTECH CO.

15

Query Match 41.4%; Score 397; DB 21; Length 832;
Best Local Similarity 45.8%; Pred. No. 5.9e-36;
Matches 82; Conservative 30; Mismatches 65; Indels 2; Gaps 2;

QY 1 FTPLAEQDCEKFAFTIPAINNKEPATRFQKWVLPQGLNSPTTCQTFVGRALQPVDRKF 60
|:||||| |:||||:|:| |:||||| |:||||| |:|:|:|
Db 113 fsplaeqdrfaftlpsvnnqparfkwvlpgqmtcsptcqlivgqilleplrikh 172

QY 61 SDCVIITHFDIICAAETKOKLIDCYTFLPAEVANAGLATASDKIQTSTPPHYLGMQJEN 120
:|| ||:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:|
Db 173 psrlmhymdmlaashdhglcaageevistieragftispdkvqrgvqylgykigs 232

QY 121 RKTKPKQIEIRKDTLTKLTDFQKLLGDINNIRPTGLTGYTAMSNLFSILRGSDSLNSKR 179
: |:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:|
Db 233 tyvapvgl-vaepriatlwdvqlvgsllwlrpalgipprlmgpfyeqlrg-sdpnear 289

RESULT 13
AAB12989
ID AAB12989 standard; Protein; 895 AA.
XX
AC AAB12989;
XX
DT 30-NOV-2000 (first entry)
XX
DE Full length avian reverse transcriptase (RT) protein.
XX
KW Reverse transcriptase; RT; polymerase chain reaction; PCR; retrovirus;
KW stability; solubility.
XX
OS Myeloblastosis-associated virus.
XX
PN WO200042199-A1.
XX
PD 20-JUL-2000.
XX
PF 14-JAN-2000; 2000WO-US00896.
XX
PI 15-JAN-1999; 99US-0116099.
XX
PA (MOLE-) MOLECULAR BIOLOGY RESOURCES.
XX
PI Swaminathan N;
XX
DR WPI; 2000-482830/42.
XX
PT Novel genes encoding reverse transcriptase polypeptides modified by
PT altering or adding the integrase domains by truncation internally
PT and/or at the C-termini, useful in cDNA synthesis and amplification
PT procedures -
XX
XX
Claim 1; Page 89-93; 189pp; English.
XX
XX This invention relates to isolated polynucleotide sequences encoding a
XX polypeptide with RNA dependent DNA polymerase (reverse transcriptase RT)
XX activity. Rns are found in a variety of retroviruses and their defining
XX activity is the ability to synthesise a cDNA strand using an RNA
XX template. The invention includes sequences AAA87808-A87822 and
XX AAA87840-A87842 which represent nucleotide sequences encoding RT
XX polypeptides. Included in the invention are reverse transcriptase
XX protein sequences AAB12989-B12995. Sequences AAA87829-A87839 and
XX AAA87843-A87897 represent oligonucleotides used in the identification and
XX synthesis of the RT nucleotide sequences of the invention. Many of the RT
XX nucleotide sequences encode modified RT proteins, which exhibit improved
XX stability and/or improved solubility, relative to naturally occurring
XX reverse transcriptases. The modified RT nucleotide and protein sequence
XX are used in improved polymerase chain reaction (PCR) methods. The
XX nucleotide sequences can be used in sequencing methods.
XX The present sequence represents a reverse transcriptase protein of the
XX invention.
XX

SQ	Sequence	895 AA;	
XX	Query Match	41.4%; Score 397; DB 21; Length 895;	
SQ	Best Local Similarity	45.8%; Pred. No. 6.6e-36;	
	Matches	82; Conservative	30; Mismatches 65; Indels 2; Gaps 2;
QY	1 FTPLAQDCCKFAFTIPAINNKPEATRFQWKVLPQGMNSPTICQTFVGRALQPVDRKF	60	
DB	113 fslplaeqrerfaftlpsvnnqparrrfqwkvlpqgmtespticqlivgileprlkh	172	
QY	61 SDCVIIHYFDDILCAETKDKLIDCYTFLPAEVANAGLAIASDKIQSTPFFHYLGMQIEN	120	
DB	173 psrlmlymddlllaashdgleaageevistleragftispdkvqrepvgvlygylgks	232	
QY	121 RKIKPQKIEKDKLTKLNDFOKLLGDINWIRPTLGIPTVAMSNIPLFSLRGDSDLNSKR	179	
DB	233 tyvapvgl-vaepriatlwdvqklvgsiqwlrpalgipprlmgpfyeqlrg-sdpnear	289	
XX	Query Match	40.4%; Score 387; DB 16; Length 895;	
SQ	Best Local Similarity	45.8%; Pred. No. 8.9e-35;	
	Matches	82; Conservative	27; Mismatches 68; Indels 2; Gaps 2;
QY	1 FTPLAQDCCKFAFTIPAINNKPEATRFQWKVLPQGMNSPTICQTFVGRALQPVDRKF	60	
DB	113 fslplaeqrerfaftlpsvnnqparrrfqwkvlpqgmtespticqlivgileprlkh	172	
QY	61 SDCVIIHYFDDILCAETKDKLIDCYTFLPAEVANAGLAIASDKIQSTPFFHYLGMQIEN	120	

Db 173 palrmihymddlllaasshdgleaagkevigtlteragftispdkigrepqvgylgyklgs 232
QY 121 RKIKPOKIEIRKDTLKTLDNDFQKLLGDINWIRPTLGIPTYAMSNLFSILRGSDSLNSKR 179
Db 233 tyvapvgi-vaepriatlwadvqklvgsiqwirpalgipprlmgpfyeqlrg-sdpnear 289

Search completed: April 9, 2002, 16:56:46
Job time: 480 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:06:34 ; Search time 70.84 Seconds
(without alignments)

375.798 Million cell updates/sec

Title: US-09-490-700-41

Perfect score: 959

Sequence: 1 FTIPLAQDCEKFAFTIPAI.....SNLFSILRGSDLSNKRMLT 182

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	949	99.0	702	12 Q9QC07	Q9qc07 human endog
2	941	98.1	956	4 Q9BXR4	Q9bxx4 homo sapien
3	941	98.1	956	4 Q9BXR3	Q9bxx3 homo sapien
4	937	97.7	596	12 Q92152	Q92152 human endog
5	937	97.7	2294	4 Q9UKH9	Q9ukh9 homo sapien
6	933	97.3	572	12 Q92154	Q92154 human endog
7	933	97.3	740	12 P87890	P87890 human endog
8	933	97.3	1361	4 Q14273	Q14273 homo sapien
9	933	97.3	1755	4 Q9UKH6	Q9ukh6 homo sapien
10	932	97.2	872	12 Q9WJH4	Q9wjh4 human endog
11	932	97.2	872	12 Q9WIK9	Q9wik9 human endog
12	932	97.2	956	4 Q9UP31	Q9up31 homo sapien
13	932	97.2	1879	4 Q9UKH5	Q9ukh5 homo sapien
14	929	96.9	572	4 Q9UQ00	Q9ugq0 homo sapien
15	924	96.4	597	12 Q92151	Q92151 human endog
16	921	96.0	875	12 Q9WJH5	Q9wjrh5 human endog
17	913	95.2	572	12 Q92153	Q92153 human endog
18	911	95.0	569	12 Q69013	Q69013 human endog
19	748	78.0	198	4 O15310	O15310 homo sapien

20	744	77.6	198	4 O15311	O15311 homo sapien
21	742	77.4	201	4 O15313	O15313 homo sapien
22	728	75.9	201	4 O15312	O15312 homo sapien
23	677	70.6	197	4 O15309	O15309 homo sapien
24	671	70.0	197	4 O15314	O15314 homo sapien
25	655	68.3	1177	4 Q9UKI1	Q9ukil homo sapien
26	537	56.0	198	12 O71090	O71090 human endog
27	531	55.4	198	12 O71089	O71089 human endog
28	525	54.7	874	12 Q9WR73	Q9wr73 sheep pulmo
29	524	54.6	198	12 O71091	O71091 human endog
30	513	53.5	766	6 Q9NIR8	Q9nir8 ovis aries
31	501	52.2	766	12 O41292	O41292 simian type
32	497	51.8	222	12 P87893	P87893 human endog
33	496	51.7	867	12 Q9WQ04	Q9wq04 simian retr
34	496	51.7	867	12 Q9WQ00	Q9wq00 simian retr
35	496	51.7	1770	12 O56224	O56224 simian maso
36	489	51.0	574	12 O41290	O41290 simian type
37	471	49.1	871	6 Q9N172	Q9n172 trichosurus
38	464	48.4	1754	12 O56220	O56220 mouse mamma
39	461	48.1	1755	12 Q91ZT3	Q91zt3 exogenous m
40	460	48.0	1755	12 Q91ZU3	Q91zt3 exogenous m
41	460	48.0	1755	12 Q91ZT8	Q91zt8 exogenous m
42	457	47.7	899	12 Q83393	Q83393 mouse mamma
43	451	47.0	714	4 Q9HBF9	Q9hbf9 homo sapien
44	395	41.2	895	12 Q04095	Q04095 avian leuko
45	395	41.2	1504	12 Q98WW2	Q98ww2 avian leuko

ALIGNMENTS

RESULT 1
Q9QC07 PRELIMINARY; PRT; 702 AA.
AC Q9QC07;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE POL. PROTEIN.
GN POL.
OS Human endogenous retrovirus K.
OC Viruses; Retroid viruses; Retroviridae.
OX NCBI_TaxID=45617;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99445825; PubMed=10516026;
RA Toenjes R.R., Czauderna F., Kurth R.;
RT "Genome wide screening, cloning, chromosomal assignment and expression
of full-length human endogenous retrovirus type K (HERV-K).";
RL J. Virol. 73:9187-9195(1999).
CC -!- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
TRANSCRIPTASE).
DR EMBL: Y18890; CAB56603.1; -;
DR InterPro: IPR003308; Integrase_zn.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF02022; Integrase_zn; 1.
DR Pfam: PF00075; rnaseh; 1.
DR Pfam: PF00665; rvt; 1.
DR Pfam: PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 702 AA; 79386 MW; D6728443636546F0 CRC64;

Query Match 99.0%; Score 949; DB 12; Length 702;

Best Local Similarity 98.9%; Pred. No. 1.7e-83;

Matches 180; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCEKFAFTIPAINNKKEPATRFQWKVLPQGLNSPTICOTFFVGRALQPVDRKF 60

|||||

DB 18 FTIPLAEQDCEKFAFTIPAINNKKEPATRFQWKVLPQGLNSPTICOTFFVGRALQPVDRKF 77

QY 61 SDYIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLAIASDKIQSTSTPFHYLGMQIEN 120
Db 78 SDYIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLAIASDKIQSTSTPFHYLGMQIEN 137
QY 121 RKIKPQKIEIRKDTLTKLNDFOKLLGDIINWIRPTLGIPTVAMSILRGDSDLNSKRM 180
Db 138 RKIKPQKIEIRKDTLTKLNDFOKLLGDIINWIRPTLGIPTVAMSILRGDSDLNSKRM 197
QY 181 LT 182
Db 198 LT 199

RESULT 2
Q9BXR4 PRELIMINARY; PRT; 956 AA.
ID Q9BXR4
AC Q9BXR4
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE POLYMERASE (FRAGMENT).
GN POL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP TISSUE=BLOOD;
RC SEQUENCE FROM N.A.
RA Reus K., Mayer J., Sauter M., Scherer D., Mueller-Lantzsch N.,
RA Meese E.;
RT "Genomic organization of the Human Endogenous Retrovirus HERV-K (HML-
2.HOM) (ERV6) on chromosome 7.";
RL Genomics 72:314-320(2001).
DR EMBL: AF298587; AAK11553.1; -.
FT NON_TER 1
SQ SEQUENCE 956 AA; 107747 MW; 648619859FC87331 CRC64;

Query Match 98.1%; Score 941; DB 4; Length 956;
Best Local Similarity 98.4%; Pred. No. 1.5e-82;
Matches 179; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 60
Db 128 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 187
QY 61 SDYIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLAIASDKIQSTSTPFHYLGMQIEN 120
Db 188 SDYIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLAIASDKIQSTSTPFHYLGMQIEN 247
QY 121 RKIKPQKIEIRKDTLTKLNDFOKLLGDIINWIRPTLGIPTVAMSILRGDSDLNSKRM 180
Db 248 RKIKPQKIEIRKDTLTKLNDFOKLLGDIINWIRPTLGIPTVAMSILRGDSDLNSKRM 307
QY 181 LT 182
Db 308 LT 309

Query Match 98.1%; Score 941; DB 4; Length 956;
Best Local Similarity 98.4%; Pred. No. 1.5e-82;
Matches 179; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 60
Db 128 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 187
QY 61 SDYIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLAIASDKIQSTSTPFHYLGMQIEN 120
Db 188 SDYIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLAIASDKIQSTSTPFHYLGMQIEN 247
QY 121 RKIKPQKIEIRKDTLTKLNDFOKLLGDIINWIRPTLGIPTVAMSILRGDSDLNSKRM 180
Db 248 RKIKPQKIEIRKDTLTKLNDFOKLLGDIINWIRPTLGIPTVAMSILRGDSDLNSKRM 307
QY 181 LT 182
Db 308 LT 309

Query Match 98.1%; Score 941; DB 4; Length 956;
Best Local Similarity 98.4%; Pred. No. 1.5e-82;
Matches 179; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

RN 1
RP SEQUENCE FROM N.A.
RC TISSUE=BLOOD;
RA Reus K., Mayer J., Sauter M., Scherer D., Mueller-Lantzsch N.,
RA Meese E.;
RT "Genomic organization of the Human Endogenous Retrovirus HERV-K (HML-
2.HOM) (ERV6) on chromosome 7.";
RL Genomics 72:314-320(2001).
DR EMBL: AF298588; AAK11554.1; -.
FT NON_TER 1
SQ SEQUENCE 956 AA; 107847 MW; 146AD510AB0F9071 CRC64;

Query Match 98.1%; Score 941; DB 4; Length 956;
Best Local Similarity 98.4%; Pred. No. 1.5e-82;
Matches 179; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 60
Db 128 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 187
QY 61 SDYIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLAIASDKIQSTSTPFHYLGMQIEN 120
Db 188 SDYIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLAIASDKIQSTSTPFHYLGMQIEN 247
QY 121 RKIKPQKIEIRKDTLTKLNDFOKLLGDIINWIRPTLGIPTVAMSILRGDSDLNSKRM 180
Db 248 RKIKPQKIEIRKDTLTKLNDFOKLLGDIINWIRPTLGIPTVAMSILRGDSDLNSKRM 307
QY 181 LT 182
Db 308 LT 309

RESULT 4
Q92152 PRELIMINARY; PRT; 596 AA.
ID Q92152
AC Q92152;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE POLYMERASE (FRAGMENT).
OS Human endogenous retrovirus K.
OC Viruses; Retroviral viruses; Retroviridae.
OX NCBI_TaxID=45617;
RN 1
RP SEQUENCE FROM N.A.
RA Berkhout B., Jebbink M., Zsiros J.;
RT "Identification of an active reverse transcriptase enzyme encoded by a
human endogenous HERV-K retrovirus.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
TRANSCRIPTASE).
DR EMBL: AF080231; AAC63291.1; -.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00075; rnaseh; 1.
DR Pfam: PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 596
SQ SEQUENCE 596 AA; 67368 MW; 59224F02A74F4980 CRC64;

Query Match 97.7%; Score 937; DB 12; Length 596;
Best Local Similarity 97.8%; Pred. No. 2e-82;
Matches 178; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 60
Db 127 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 186
QY 61 SDYIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLAIASDKIQSTSTPFHYLGMQIEN 120

Query Match 97.7%; Score 937; DB 12; Length 596;
Best Local Similarity 97.8%; Pred. No. 2e-82;
Matches 178; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 60
Db 127 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 186
QY 61 SDYIIHYFDDILCAAEKDKLIDCYTFLPAEVANAGLAIASDKIQSTSTPFHYLGMQIEN 120

```
Db 187 SDYIIHYDDILCAAETKDKLIDCYTFLOAEVANAGLAIASDKIQTSTPFHYLGMQIEN 246
|||
QY 121 RKIKPQKIEIRKDTLTKLNDFOKLLGDINWIRPTLGIPTIYAMNLSILRGSDSLNSKRM 180
|||
Db 247 RKIKPQKIEIRKDTLTKLNDFOKLLGDINWIRPTLGIPTIYAMNLSILRGSDSLNSKRI 306
|||
QY 181 LT 182
||
Db 307 LT 308

RESULT 5
Q9UKH9 ID Q9UKH9 PRELIMINARY; PRT: 2294 AA.
AC Q9UKH9;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DE GAG-PRO-POL-ENV PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99400989; PubMed=10469592;
RA Barbulescu M., Turner G., Seaman M.I., Deinard A.S., Kidd K.K.,
RA Lenz J.;
RT "Many human endogenous retrovirus K (HERV-K) proviruses are unique to
RT humans.";
RL Curr. Biol. 9:861-868(1999).
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
CC TRANSCRIPTASE).
CC -1- SIMILARITY: BELONGS TO ZN-FINGER CCHC TYPE FAMILY.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO THE PEPTIDASE FAMILY A1; ALSO
CC KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR EMBL: AF164611; AADS1793.1; -.
DR HSSP: P16088; 1DUT.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001995; Asp_prot_retrov.
DR InterPro: IPR001428; dUTPase.
DR InterPro: IPR003322; Gag_p10.
DR InterPro: IPR000721; Gag_p24.
DR InterPro: IPR000467; G_patch.
DR InterPro: IPR001037; Integrase_C.
DR InterPro: IPR003308; Integrase_zn.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RVTse.
DR InterPro: IPR001878; znF_CCHC.
DR Pfam: PF00592; dUTPase; 1.
DR Pfam: PF01585; G_patch; 1.
DR Pfam: PF02337; Gag_p10; 1.
DR Pfam: PF06077; Gag_p10; 1.
DR Pfam: PF06077; gag_p24; 1.
DR Pfam: PF00552; Integrase; 1.
DR Pfam: PF02022; Integrase_zn; 1.
DR Pfam: PF00075; rnaseh; 1.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00077; rvp; 1.
DR Pfam: PF00078; rvt; 1.
DR Pfam: PF00098; zf-CCHC; 2.
DR ProDom: PD000946; dUTPase; 1.
DR SMART: SM00443; G_patch; 1.
DR SMART: SM00343; znF_C2HC; 2.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
DR PROSITE: PS50175; ASP_PROT_RETROV; 1.
KW Aspartyl protease; Hydrolase; RNA-directed DNA polymerase;
KW Zinc-finger.
SQ SEQUENCE 2294 AA; 257099 MW; 425C81FDDF8A80D0 CRC64;
```

Query Match 97.7%; Score 937; DB 4; Length 2294;

```
Best Local Similarity 97.8%; Pred. No. 1.le-81;
Matches 178; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVKDF 60
|||||
Db 1051 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 1110
|||||
QY 61 SDYIIHYFDDILCAAETKDKLIDCYTFLOAEVANAGLAIASDKIQTSTPFHYLGMQIEN 120
|||||
Db 1111 SDYIIHYDDILCAAETKDKLIDCYTFLOAEVANAGLAIASDKIQTSTPFHYLGMQIEN 1170
|||||
QY 121 RKIKPQKIEIRKDTLTKLNDFOKLLGDINWIRPTLGIPTIYAMNLSILRGSDSLNSKRM 180
|||||
Db 1171 RKIKPQKIEIRKDTLTKLNDFOKLLGDINWIRPTLGIPTIYAMNLSILRGSDSLNSKRI 1230
|||||
QY 181 LT 182
||
Db 1231 LT 1232

RESULT 6
Q92154 ID Q92154 PRELIMINARY; PRT: 572 AA.
AC Q92154;
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DE POLYMERASE (FRAGMENT).
OS Human endogenous retrovirus K.
OC Viruses; Retroid viruses; Retroviridae.
OX NCBI_TaxID=45617;
RN [1]
RP SEQUENCE FROM N.A.
RA Berkhout B., Jebbink M., Zsiros J.;
RT "Identification of an active reverse transcriptase enzyme encoded by a
RT human endogenous HERV-K retrovirus.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
CC TRANSCRIPTASE).
DR EMBL: AF080234; AAC63294.1; -.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00075; rnaseh; 1.
DR Pfam: PF00078; rvt; 1.
DR RNA-directed DNA polymerase.
KW NON_TER 1
FT NON_TER 572
SQ SEQUENCE 572 AA; 64753 MW; FDD2AFA37A3A7D0 CRC64;

Query Match 97.3%; Score 933; DB 12; Length 572;
Best Local Similarity 97.8%; Pred. No. 4.7e-82;
Matches 178; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVKDF 60
|||||
Db 111 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 170
|||||
QY 61 SDYIIHYFDDILCAAETKDKLIDCYTFLOAEVANAGLAIASDKIQTSTPFHYLGMQIEN 120
|||||
Db 171 SDYIIHYDDILCAAETKDKLIDCYTFLOAEVANAGLAIASDKIQTSTPFHYLGMQIEN 230
|||||
QY 121 RKIKPQKIEIRKDTLTKLNDFOKLLGDINWIRPTLGIPTIYAMNLSILRGSDSLNSKRM 180
|||||
Db 231 RKIKPQKIEIRKDTLTKLNDFOKLLGDINWIRPTLGIPTIYAMNLSILRGSDSLNSKRM 290
|||||
QY 181 LT 182
||
Db 291 LT 292

RESULT 7
PB7890
```

```

ID P87890 PRELIMINARY; PRT; 740 AA.
AC P87890;
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE POL PROTEIN (FRAGMENT).
GN POL.
OS Human endogenous retrovirus K.
OC Viruses; Retroviral viruses; Retroviridae.
OX NCBI_TaxID=45617;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT GH.
RX MEDLINE=97360024; PubMed=9217052;
RA Toenjes R.R., Boller K., Limbach R., Lugert R., Kurth R.;
RT "Characterization of human endogenous retrovirus type K virus-like
particles generated from recombinant baculoviruses.";
RL Virology 233:280-291(1997).
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
TRANSCRIPTASE).
DR EMBL; Y10391; CAA71417.1; -.
DR InterPro: IPR003308; Integrase_zn.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RVtse.
DR Pfam; PF02022; Integrase_zn; 1.
DR Pfam; PF00075; rnaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
FT NON_TER 740 740
FT NON_TER 740 740
SQ SEQUENCE 740 AA; 83564 MW; 76AD8F9EA35B6BDB CRC64;

Query Match 97.3%; Score 933; DB 12; Length 740;
Best Local Similarity 97.3%; Pred. No. 6.5e-82;
Matches 177; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 FTIPLAQDCCKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 60
DB 128 FTIPLAQDCCKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 187
QY 61 SDCVIIHYFDLILCAAEKDKLIDCYTFLPAEVANAGLATASDKIQSTPFFHYLGMQIEN 120
DB 188 SDCVIIHYIDDILCAAEKDKLIDCYTFLQAEVANAGLATASDKIQSTPFFHYLGMQIEN 247
QY 121 RKIKPKQIEIRKDKTLNDFQKLLGDIINWIRPTLGIPTYAMS NFLSILRGDSLSNKR 180
DB 248 RKIKPKQIEIRKDKTLNDFQKLLGDIINWIRPTLGIPTYAMS NFLSILRGDSLSNQR 307
QY 181 LT 182
DB 308 LT 309

RESULT 8
Q14273 PRELIMINARY; PRT; 1361 AA.
AC Q14273;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE POL/ENV ORF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=87036922; PubMed=3021993;
RA Ono M., Yasunaga T., Miyata T., Ushikubo H.;

ID Q14273 PRELIMINARY; PRT; 1755 AA.
AC Q14273;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE GAG-PRO-POL PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99400989; PubMed=10469592;
RA Barbulescu M., Turner G., Seaman M.I., Deinard A.S., Kidd K.K.,
RA Lenz J.;
RT "Many human endogenous retrovirus K (HERV-K) proviruses are unique to
humans.";
RL Curr. Biol. 9:861-868(1999).
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
TRANSCRIPTASE).
CC -1- SIMILARITY: BELONGS TO ZN-FINGER CCHC TYPE FAMILY.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO THE PEPTIDASE FAMILY A1; ALSO
KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR EMBL; AF164613; AAD51796.1; -.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001995; Asp_prot_retrov.
DR InterPro: IPR001428; dutPase.
DR InterPro: IPR003322; Gag_p10.
DR InterPro: IPR000721; Gag_p24.
DR InterPro: IPR000467; G_patch.

```

```

RT "Nucleotide sequence of human endogenous retrovirus genome related to
the mouse mammary tumor virus genome.";
RL J. Virol. 60:589-598(1986).
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
TRANSCRIPTASE).
DR EMBL; M14123; AAA88033.1; -.
DR InterPro: IPR001037; Integrase_C.
DR InterPro: IPR003308; Integrase_zn.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RVtse.
DR Pfam; PF00552; Integrase_zn; 1.
DR Pfam; PF02022; Integrase_zn; 1.
DR Pfam; PF00075; rnaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1361 AA; 153797 MW; CEB91B3F407B9498 CRC64;

Query Match 97.3%; Score 933; DB 4; Length 1361;
Best Local Similarity 97.3%; Pred. No. 1.4e-81;
Matches 177; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 FTIPLAQDCCKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 60
DB 30 FTIPLAQDCCKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 89
QY 61 SDCVIIHYFDLILCAAEKDKLIDCYTFLPAEVANAGLATASDKIQSTPFFHYLGMQIEN 120
DB 90 SDCVIIHYIDDILCAAEKDKLIDCYTFLQAEVANAGLATASDKIQSTPFFHYLGMQIEN 149
QY 121 RKIKPKQIEIRKDKTLNDFQKLLGDIINWIRPTLGIPTYAMS NFLSILRGDSLSNKR 180
DB 150 RKIKPKQIEIRKDKTLNDFQKLLGDIINWIRPTLGIPTYAMS NFLSILRGDSLSNQR 209
QY 181 LT 182
DB 210 LT 211

RESULT 9
Q9UKH6 PRELIMINARY; PRT; 1755 AA.
AC Q9UKH6;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE GAG-PRO-POL PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99400989; PubMed=10469592;
RA Barbulescu M., Turner G., Seaman M.I., Deinard A.S., Kidd K.K.,
RA Lenz J.;
RT "Many human endogenous retrovirus K (HERV-K) proviruses are unique to
humans.";
RL Curr. Biol. 9:861-868(1999).
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
TRANSCRIPTASE).
CC -1- SIMILARITY: BELONGS TO ZN-FINGER CCHC TYPE FAMILY.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO THE PEPTIDASE FAMILY A1; ALSO
KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR EMBL; AF164613; AAD51796.1; -.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001995; Asp_prot_retrov.
DR InterPro: IPR001428; dutPase.
DR InterPro: IPR003322; Gag_p10.
DR InterPro: IPR000721; Gag_p24.
DR InterPro: IPR000467; G_patch.

```



```
DR InterPro: IPR003308; Integrase_zn.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RVTse.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00692; dnpase; 1.
DR Pfam: PF01385; G-patch; 1.
DR Pfam: PF02337; Gag_p10; 1.
DR Pfam: PF00607; gag_p24; 1.
DR Pfam: PF02022; Integrase_zn; 1.
DR Pfam: PF00075; rnaseh; 1.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00077; rvp; 1.
DR Pfam: PF00078; rvt; 1.
DR Pfam: PF00098; zf-CCHC; 2.
DR ProDom: PD000946; dnpase; 1.
DR SMART: SM00443; G-patch; 1.
DR SMART: SM00343; Znf_C2HC; 2.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
DR PROSITE: PS0175; ASP_PROT_RETROV; 1.
KW Aspartyl protease; Hydrolase; RNA-directed DNA polymerase;
KW Zinc-finger.
SQ SEQUENCE 1755 AA; 195683 MW; DC348F16E45687BC CRC64;

Query Match 97.3%; Score 933; DB 4; Length 1755;
Best Local Similarity 97.3%; Pred. No. 1.9e-81;
Matches 177; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 60
Db 1051 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 1110
QY 61 SDCYIIHYFDDILCAAEKDKLDCYTFELPAEVANAGLAIASDKIQSTPFFHYLGMQIEN 120
Db 1111 SDCYIIHYFDDILCAAEKDKLDCYTFELPAEVANAGLAIASDKIQSTPFFHYLGMQIEN 1170
QY 121 RKIKPQKIEIRKDTLKTLDNFQKLLGDINWIRPTLGIPTVYAMSNLFSILRGSDSLNSKRM 180
Db 1171 RKIKPQKIEIRKDTLKTLDNFQKLLGDINWIRPTLGIPTVYAMSNLFSILRGSDSLNSQRI 1230
QY 181 LT 182
Db 1231 LT 1232

RESULT 10
Q9WJR4 PRELIMINARY; PRT; 872 AA.
AC Q9WJR4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE POL PROTEIN.
GN GAG-POL OR POL.
OS Human endogenous retrovirus K.
OC Viruses; Retroid viruses; Retroviridae.
OX NCBI_TaxID=45617;
RN [1]
RP SEQUENCE FROM N.A.
RA Toenjes R.R., Czauderna F., Kurth R.;
RT "Full-length human endogenous retrovirus type K (HERV-K) elements encoding Gag, Pol and Env proteins are localised on chromosomes 7 and 19."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- PTM: SPECIFIC ENZYMTIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS (BY SIMILARITY).
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE TRANSCRIPTASE).
CC -1- SIMILARITY: TO RETROVIRAL INTEGRASE, C-TERMINUS.
DR EMBL; Y17834; CAA76885.1; -.
DR InterPro: IPR001037; Integrase_C.
DR InterPro: IPR003308; Integrase_zn.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00552; Integrase; 1.
DR Pfam: PF02022; Integrase; 1.
DR Pfam: PF02022; Integrase; 1.
DR Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- PTM: SPECIFIC ENZYMTIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS (BY SIMILARITY).
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE TRANSCRIPTASE).
CC -1- SIMILARITY: TO RETROVIRAL INTEGRASE, C-TERMINUS.
DR EMBL; Y17834; CAA76885.1; -.
DR InterPro: IPR001037; Integrase_C.
DR InterPro: IPR003308; Integrase_zn.
```

```
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00552; Integrase; 1.
DR Pfam: PF02022; Integrase_zn; 1.
DR Pfam: PF00075; rnaseh; 1.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00078; rvt; 1.
DR Pfam: PF00077; rvp; 1.
KW DNA-binding; Endonuclease; Hydrolase; Nucleotidyltransferase;
KW Polyprotein; RNA-directed DNA polymerase.
SQ SEQUENCE 872 AA; 97873 MW; FAA4267E3B5B8C8B CRC64;

Query Match 97.2%; Score 932; DB 12; Length 872;
Best Local Similarity 97.8%; Pred. No. 1e-81;
Matches 178; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 60
Db 44 FTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGLNSPTICQTFVGRALQPVREKF 103
QY 61 SDCYIIHYFDDILCAAEKDKLDCYTFELPAEVANAGLAIASDKIQSTPFFHYLGMQIEN 120
Db 104 SDCYIIHYFDDILCAAEKDKLDCYTFELPAEVANAGLAIASDKIQSTPFFHYLGMQIEN 163
QY 121 RKIKPQKIEIRKDTLKTLDNFQKLLGDINWIRPTLGIPTVYAMSNLFSILRGSDSLNSKRM 180
Db 164 RKIKPQKIEIRKDTLKTLDNFQKLLGDINWIRPTLGIPTVYAMSNLFSILRGSDSLNSKRM 223
QY 181 LT 182
Db 224 LT 225

RESULT 11
Q9WIK9 PRELIMINARY; PRT; 872 AA.
AC Q9WIK9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE POL PROTEIN.
GN GAG-POL OR POL.
OS Human endogenous retrovirus K.
OC Viruses; Retroid viruses; Retroviridae.
OX NCBI_TaxID=45617;
RN [1]
RP SEQUENCE FROM N.A.
RA Toenjes R.R., Czauderna F., Kurth R.;
RT "Full-length human endogenous retrovirus type K (HERV-K) elements encoding Gag, Pol and Env proteins are localised on chromosomes 7 and 19."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- PTM: SPECIFIC ENZYMTIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS (BY SIMILARITY).
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE TRANSCRIPTASE).
CC -1- SIMILARITY: TO RETROVIRAL INTEGRASE, C-TERMINUS.
DR EMBL; Y17832; CAA76879.1; -.
DR InterPro: IPR001037; Integrase_C.
DR InterPro: IPR003308; Integrase_zn.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00552; Integrase; 1.
DR Pfam: PF02022; Integrase; 1.
DR Pfam: PF00075; rnaseh; 1.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00078; rvt; 1.
KW DNA-binding; Endonuclease; Hydrolase; Nucleotidyltransferase;
KW Polyprotein; RNA-directed DNA polymerase.
SQ SEQUENCE 872 AA; 97876 MW; E5D0A2390060BD5D CRC64;
```

```
Query Match          97.2%; Score 932; DB 12; Length 872;
Best Local Similarity 97.8%; Pred. No. 1e-81;
Matches 178; Conservative 1; Mismatches 0; Gaps 0;

Qy 1 FTIPLAEODCEKFAFTIPAINNKEPATRFQWKVLPQGMNSPTTCQTFVGRALQPVREKF 60
Db 128 FTIPLAEODCEKFAFTIPAINNKEPATRFQWKVLPQGMNSPTTCQTFVGRALQPVREKF 187

Qy 61 SDCTIIHVFDDILCAAEKDKLIDCYTFLPAEVANAGIAIASDKIQSTPPTHYLGQMQIEN 120
Db 188 SDCTIIHVFDDILCAAEKDKLIDCYTFLPAEVANAGIAIASDKIQSTPPTHYLGQMQIEN 247

Qy 121 RKIPQKIEIRKDTLKTNDLNDQKLLGDNWIRPTLGIPTIYAMSNLFSILRGSDSLNSKRM 180
Db 248 RKIPQKIEIRKDTLKTNDLNDQKLLGDNWIRPTLGIPTIYAMSNLFSILRGSDSLNSKRM 307

Qy 181 LT 182
Db 308 LT 309

RESULT 13
Q9UKH5 ID Q9UKH5 PRELIMINARY; PRT; 1879 AA.
AC Q9UKH5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GAG-PRO-POL PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99400983; PubMed=10469592;
RA Barbulescu M., Turner G., Seaman M.I., Deinard A.S., Kidd K.K.,
RA Lenz J.;
RT "Many human endogenous retrovirus K (HERV-K) proviruses are unique to
RT humans.";
RL Curr. Biol. 9:861-868(1999).
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
CC TRANSCRIPTASE).
CC -1- SIMILARITY: BELONGS TO ZN-FINGER CCHC TYPE FAMILY.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO THE PEPTIDASE FAMILY A1; ALSO
CC KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR EMBL; AF164614; AAD51797.1; -.
DR InterPro; IPR001969; Asp.protease.
DR InterPro; IPR001995; Asp_prot_retrov.
DR InterPro; IPR001428; dUTPase.
DR InterPro; IPR003322; Gag_p10.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR000467; G_patch.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_zn.
DR InterPro; IPR001584; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF006932; dUTPase; 1.
DR Pfam; PF01585; G_patch; 1.
DR Pfam; PF02337; Gag_p10; 1.
DR Pfam; PF00607; gag_p24; 1.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF02022; Integrase_zn; 1.
DR Pfam; PF00075; rnaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; rvp; 1.
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF00098; zf-CCHC; 2.
DR ProDom; PD000946; dUTPase; 1.
DR SMART; SM00443; G_patch; 1.
DR SMART; SM0343; Znf_C2HC; 2.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS01075; ASP_PROT_RETROV; 1.
KW Aspartyl protease; Hydrolase; RNA-directed DNA polymerase;
KW Zinc-finger.
```

```
Query Match          97.2%; Score 932; DB 12; Length 872;
Best Local Similarity 97.8%; Pred. No. 1e-81;
Matches 178; Conservative 1; Mismatches 0; Gaps 0;

Qy 1 FTIPLAEODCEKFAFTIPAINNKEPATRFQWKVLPQGMNSPTTCQTFVGRALQPVREKF 60
Db 44 FTIPLAEODCEKFAFTIPAINNKEPATRFQWKVLPQGMNSPTTCQTFVGRALQPVREKF 103

Qy 61 SDCTIIHVFDDILCAAEKDKLIDCYTFLPAEVANAGIAIASDKIQSTPPTHYLGQMQIEN 120
Db 104 SDCTIIHVFDDILCAAEKDKLIDCYTFLPAEVANAGIAIASDKIQSTPPTHYLGQMQIEN 163

Qy 121 RKIPQKIEIRKDTLKTNDLNDQKLLGDNWIRPTLGIPTIYAMSNLFSILRGSDSLNSKRM 180
Db 164 RKIPQKIEIRKDTLKTNDLNDQKLLGDNWIRPTLGIPTIYAMSNLFSILRGSDSLNSKRM 223

Qy 181 LT 182
Db 224 LT 225

RESULT 12
Q9UP31 ID Q9UP31 PRELIMINARY; PRT; 956 AA.
AC Q9UP31;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE POLYMERASE (FRAGMENT).
GN POL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99178257; PubMed=10080172;
RA Mayer J., Sauter M., Racz A., Scherer D., Mueller-Lantzsch N.,
RA Meese E.;
RT "An almost-intact human endogenous retrovirus K on human chromosome
RT 7.";
RL Nat. Genet. 21:257-258(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Reus K., Mayer J., Sauter M., Scherer D., Mueller-Lantzsch N.,
RA Meese E.;
RT "Further characterization of the almost intact human endogenous
RT retrovirus K on human chromosome 7.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
CC TRANSCRIPTASE).
DR EMBL; AF074086; AAD21097.1; -.
DR EMBL; AF074086; AAF88167.1; -.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_zn.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF02022; Integrase_zn; 1.
DR Pfam; PF00075; rnaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00078; rvt; 1.
DR RNA-directed DNA polymerase.
KW NON_TER
SQ SEQUENCE 956 AA; 107687 MW; E6872987ADCAE41A CRC64;
```

```
Query Match          97.2%; Score 932; DB 4; Length 956;
Best Local Similarity 97.8%; Pred. No. 1.1e-81;
Matches 178; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
SQ SEQUENCE 1879 AA; 208995 MW; F95204AA2E3B10AD CRC64;

Query Match          97.2%; Score 932; DB 4; Length 1879;
Best Local Similarity 97.8%; Pred. NO. 2.6e-81;
Matches 178; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCEKEFAFTIPAINNKEPATRFQWKVLPQGMNSPTICQTFVGRALQPVREKDF 60
DB 1051 FTIPLAEQDCEKEFAFTIPAINNKEPATRFQWKVLPQGMNSPTICQTFVGRALQPVREKDF 1110
QY 61 SDCYIIHYFDDILCAAETKDKLDCYTFLLPAEVANAGLAIASDKIQSTPFFHYLGMQIEN 120
DB 1111 SDCYIIHYFDDILCAAETKDKLDCYTFLLPAEVANAGLAIASDKIQSTPFFHYLGMQIEN 1170
QY 121 RKIKPKQIEIRKDTLTLNDFQKLLGDINWIRPTLGIPTIYAMNSLFSILRGSDSLNSKRM 180
DB 1171 RKIKPKQIEIRKDTLTLNDFQKLLGDINWIRPTLGIPTIYAMNSLFSILRGSDSLNSKRM 1230
QY 181 LT 182
DB 1231 LT 1232

RESULT 14
QYUQGO PRELIMINARY; PRT; 572 AA.
AC QYUQGO;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE POLYMERASE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RX MEDLINE=99139020; PubMed=9971820;
RA Berkhout B., Jebbink M., Zsiros J.;
RT "Identification of an active reverse transcriptase enzyme encoded by a
human endogenous HERV-K retrovirus.";
RL J. Virol. 73:2365-2375(1999).
CC -!- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
TRANSCRIPTASE).
DB EMBL: AF080232; AAC63292.1; -.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00075; rNaseH; 1.
DR Pfam: PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 572
SQ SEQUENCE 572 AA; 64684 MW; A431D47DCE09C64F CRC64;

Query Match          96.9%; Score 929; DB 4; Length 572;
Best Local Similarity 97.3%; Pred. NO. 1.2e-81;
Matches 177; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCEKEFAFTIPAINNKEPATRFQWKVLPQGMNSPTICQTFVGRALQPVREKDF 60
DB 111 FTIPLAEQDCEKEFAFTIPAINNKEPATRFQWKVLPQGMNSPTICQTFVGRALQPVREKDF 170
QY 61 SDCYIIHYFDDILCAAETKDKLDCYTFLLPAEVANAGLAIASDKIQSTPFFHYLGMQIEN 120
DB 171 SDCYIIHYFDDILCAAETKDKLDCYTFLLPAEVANAGLAIASDKIQSTPFFHYLGMQIEN 230
QY 121 RKIKPKQIEIRKDTLTLNDFQKLLGDINWIRPTLGIPTIYAMNSLFSILRGSDSLNSKRM 180
DB 231 RKIKPKQIEIRKDTLTLNDFQKLLGDINWIRPTLGIPTIYAMNSLFSILRGSDSLNSKRI 290
```

```
QY 181 LT 182
DB 291 LT 292

RESULT 15
QYUQGO PRELIMINARY; PRT; 597 AA.
AC QYUQGO;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE POLYMERASE (FRAGMENT).
OS Human endogenous retrovirus K.
OC Viruses; Retroviral viruses; Retroviridae.
OX NCBI_TaxID=45617;
RN [1]
RP SEQUENCE FROM N.A.
RA Berkhout B., Jebbink M., Zsiros J.;
RT "Identification of an active reverse transcriptase enzyme encoded by a
human endogenous HERV-K retrovirus.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
TRANSCRIPTASE).
DB EMBL: AF080229; AAC63290.1; -.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00075; rNaseH; 1.
DR Pfam: PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 597
SQ SEQUENCE 597 AA; 67446 MW; B5D80527477E0FEC CRC64;

Query Match          96.4%; Score 924; DB 12; Length 597;
Best Local Similarity 95.6%; Pred. NO. 3.7e-81;
Matches 174; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 FTIPLAEQDCEKEFAFTIPAINNKEPATRFQWKVLPQGMNSPTICQTFVGRALQPVREKDF 60
DB 127 FTIPLAEQDCEKEFAFTIPAINNKEPATRFQWKVLPQGMNSPTICQTFVGRALQPVREKDF 186
QY 61 SDCYIIHYFDDILCAAETKDKLDCYTFLLPAEVANAGLAIASDKIQSTPFFHYLGMQIEN 120
DB 187 SDCYIIHYFDDILCAAETKDKLDCYTFLLPAEVANAGLAIASDKIQSTPFFHYLGMQIEN 246
QY 121 RKIKPKQIEIRKDTLTLNDFQKLLGDINWIRPTLGIPTIYAMNSLFSILRGSDSLNSKRM 180
DB 247 RKIKPKQIEIRKDTLTLNDFQKLLGDINWIRPTLGIPTIYAMNSLFSILRGSDSLNSKRI 306
QY 181 LT 182
DB 307 LT 308

Search completed: April 9, 2002, 17:06:35
Job time: 584 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 16:56:44 ; Search time 72.35 Seconds
(without alignments)
185.311 Million cell updates/sec

Title: US-09-490-700-40

Perfect score: 978

Sequence: 1 MVTPTWMDNPIEVVDSV.....POLLRTNSWSKCTSESSCR 181

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*

- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	978	100.0	181	20	AAW95694 Human endogenous r
2	978	100.0	181	20	AAW97747 Human endogenous r
3	840	85.9	153	20	AAW95692 Human endogenous r
4	840	85.9	153	20	AAW97745 Human endogenous r
5	840	85.9	561	20	AAW95693 Human endogenous r
6	840	85.9	561	20	AAW97746 Human endogenous r
7	220	22.5	48	22	AAW14052 Peptide #486 encod
8	220	22.5	48	22	AAW26458 Peptide #495 encod
9	220	22.5	48	22	AAW01792 Peptide #474 encod
10	80.5	8.2	1003	13	AAW29648 AmEPV Spheroidin p
11	80.5	8.2	1003	15	AAW55576 AmEPV Spheroidin.

12	80.5	8.2	1003	19	AAW41301 AmEPV entomopoxvir
13	80.5	8.2	1003	20	AAW30169 Spheroidin protein
14	80.5	8.2	2813	19	AAW54347 Canine von Willebr
15	80.5	8.2	2813	21	AAV70557 Canine von Willebr
16	79	8.1	191	21	AAW08013 Arabidopsis thalia
17	79	8.1	191	21	AAW43910 Arabidopsis thalia
18	79	8.1	295	21	AAW08012 Arabidopsis thalia
19	79	8.1	295	21	AAW43909 Arabidopsis thalia
20	79	8.1	305	21	AAW43908 Arabidopsis thalia
21	79	8.1	306	21	AAW08011 Arabidopsis thalia
22	76	7.8	488	22	AAU14647 Novel bone marrow
23	72.5	7.4	687	22	AAW97728 NSI protein sequen
24	72.5	7.4	1420	18	AAW13826 yeast transcriptio
25	72.5	7.4	1475	11	AAW08221 Recombinant alpha
26	72	7.4	688	20	AAW41142 Mouse mammary tumo
27	71.5	7.3	335	22	AAW06581 Human protein havi
28	71	7.3	134	21	AAW55721 Arabidopsis thalia
29	71	7.3	149	21	AAW55720 Arabidopsis thalia
30	71	7.3	158	21	AAW55719 Arabidopsis thalia
31	71	7.3	185	17	AAW03547 ORF-4 protein sequ
32	71	7.3	725	22	AAW72914 E. coli ironNec ext
33	70.5	7.2	410	20	AAW49062 Mouse membrane dip
34	70.5	7.2	1248	22	AAW39055 Human polypeptide
35	70	7.2	252	21	AAW47392 Arabidopsis thalia
36	70	7.2	256	21	AAW31623 Arabidopsis thalia
37	70	7.2	260	21	AAW44355 Arabidopsis thalia
38	69.5	7.1	1464	17	AAW88469 Feline infectious
39	69	7.1	185	21	AAW31812 Arabidopsis thalia
40	69	7.1	196	21	AAW31811 Arabidopsis thalia
41	69	7.1	204	21	AAW31810 Arabidopsis thalia
42	69	7.1	468	19	AAW74797 Human secreted pro
43	69	7.1	549	20	AAW60008 Human endometrium
44	69	7.1	1086	20	AAW81839 Human LFA-1 alpha
45	69	7.1	1502	22	AAW39273 Human polypeptide

ALIGNMENTS

RESULT 1
AAW95694
ID AAW95694 standard; Protein; 181 AA.
XX
AC AAW95694:
DT 08-JUN-1999 (first entry)
XX
DE Human endogenous retrovirus IDMK1.2-22 env/fs (sag) protein.
XX
KW IDDM; insulin-dependent diabetes mellitus; endogenous retrovirus;
KW SAg; superantigen; provirus; autoimmune disease; type 1 diabetes;
KW diagnosis; env; envelope.
XX
OS Homo sapiens.
XX
PN EP893691-A1.
XX
PD 27-JAN-1999.
XX
PF 23-JUL-1997; 97EP-0401773.
XX
PR 23-JUL-1997; 97EP-0401773.
XX
PA (MACH/) MACH B F.
XX
PI Conrad B, Mach B;
XX
DR WPI; 1999-097928/09.
XX
DR N-PSDB; AAX07516.
XX
PT Diagnosing human autoimmune disease by detecting retrovirus with
PT superantigen activity - new retrovirus associated with type 1
PT diabetes, its proviral DNA, and related vectors, transformed cells,

PT proteins, antibodies and specific binding agents, used for treating
PT or preventing autoimmune disease
PS Claim 31; Fig 7G; 92pp; English.
XX
XX The sequence is that of an insulin-dependent diabetes mellitus
XX associated human endogenous retrovirus (IDMK1.2-22) env/fs protein.
XX The retrovirus has superantigen (SAG) activity. It can be used
XX as part of a method is specifically used to diagnose type 1 diabetes
XX mellitus. Modified proteins expressed by the retroviral sequence
XX (without SAG activity but still able to induce an immune response)
XX are useful in vaccines to treat or prevent SAG-related autoimmune
XX disease; nucleic acid sequences encoding (modified) SAG can be used
XX similarly to treat such diseases. Retroviral-encoded SAG are important
XX in pathogenesis of autoimmune disease, probably by activating
XX autoreactive T cells. The method is very specific (it can differentiate
XX between expressed and non-expressed viral nucleic acids) and can be used
XX even where the pathogen is an ubiquitous endogenous retrovirus. Blood
XX or plasma samples can be tested without extensive preparation and
XX diagnosis can be made before clinical signs are apparent, allowing
XX early intervention before severe tissue damage has occurred.
XX
XX Sequence 181 AA;

Query Match 100.0%; Score 978; DB 20; Length 181;
Best Local Similarity 100.0%; Pred. No. 6.7e-106;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVTPTWMDNPIEVYVNDVSVVPGPTDDRCAPKPEEGMMINISIGYHYPPICIGRAPGC 60
Db 1 mvtptwmdnpienvyvdsvvpgptddrcapkeegmmminisigyhyppicigrpgc 60
Qy 61 LMPAVQNLVVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYORSKLFKPKGKTCPEI 120
Db 61 lmpavqnlvvevptvspnsrftyhmvsgrmlrprvnylqdfsyqrsklfrpkgtcpei 120
Qy 121 PKGSKNTEVLWEECVANSVILQNNFEFTIIDLGTSRSLPQLLRNNSVVSCKTSSSC 180
Db 121 pkgskntevlwecvansvvlqnnfeftiidlgtsrslpqlrrnnsvvscktsessc 180
Qy 181 R 181
Db 181 r 181

RESULT 2
AAW97747
ID AAW97747 standard; Protein; 181 AA.
XX
AC AAW97747;

XX 21-MAY-1999 (first entry)
XX Human endogenous retrovirus IDDK1.2-22 Env/F-S (SAG).

XX HERV; IDDK1.2-22; superantigen; SAG; antigen; IDDM;
KW insulin-dependent diabetes mellitus; autoimmune disease; diagnosis;
KW therapy; vaccine; envelope protein; env gene.
XX Human endogenous retrovirus.

XX Key Location/Qualifiers
FH Modified-site 17..19
FT /note= "Asn is N-glycosylated"
FT Modified-site 42..45
FT /note= "Asn is N-glycosylated"
XX

XX WO9905527-A2.

XX 04-FEB-1999.

XX 22-JUL-1998; 98WO-EP04926.

XX 23-JUL-1997; 97EP-0401773.
PR 22-JUL-1997; 97EP-0112482.
XX
XX (MEDI-) MEDIGEN SA.
XX
XX Conrad B, Mach B;
XX
XX WPI: 1999-143118/12.
DR N-PSDB; AAX07191.
XX
XX New isolated human endogenous retrovirus - used to develop products
PT for the diagnosis, prevention and treatment of autoimmune disease,
PT particularly insulin dependent diabetes mellitus
XX
XX Claim 6; Fig 7G; 165pp; English.

XX This is the envelope (Env)/FS polypeptide of a new human endogenous
CC retrovirus (HERV), designated IDDMK1.2-22, that has been identified
CC as the source of superantigen (SAG) activity in insulin-dependent
CC diabetes mellitus (IDDM) patients. The endogenous retrovirus is
CC ubiquitous in the human genome but is only expressed in diabetic
CC individuals. The HERV encodes SAG activity within the env gene.
CC A claimed process for the diagnosis, including the pre-symptomatic
CC diagnosis, of a human autoimmune disease associated with a HERV
CC having SAG activity comprises specifically detecting in a
CC biological sample either: (a) the mRNA of an expressed HERV having
CC SAG activity (especially IDDMK1.2-22 5'LTR, 3' long terminal
CC repeat, env or pol); (b) a protein or peptide expressed by the HERV
CC (see AAW97745-48); (c) antibodies specific to the proteins expressed
CC by the HERV; or (d) SAG activity specifically associated with the
CC HERV. Products of the invention can be used to identify substances
CC capable of blocking transcription or translation of SAG-encoding
CC nucleic acid sequences, useful in therapy and/or prevention of
CC autoimmune disease associated with the SAG. A nucleic acid encoding
CC human retroviral SAG can be used as a DNA vaccine. Expression of
CC the endogenous SAG in IDDM suggests a general model according to
CC which self SAG-driven and systemic activation of autoreactive T
CC cells leads to organ-specific autoimmune disease.

XX Sequence 181 AA;

Query Match 100.0%; Score 978; DB 20; Length 181;
Best Local Similarity 100.0%; Pred. No. 6.7e-106;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVTPTWMDNPIEVYVNDVSVVPGPTDDRCAPKPEEGMMINISIGYHYPPICIGRAPGC 60
Db 1 mvtptwmdnpienvyvdsvvpgptddrcapkeegmmminisigyhyppicigrpgc 60
Qy 61 LMPAVQNLVVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYORSKLFKPKGKTCPEI 120
Db 61 lmpavqnlvvevptvspnsrftyhmvsgrmlrprvnylqdfsyqrsklfrpkgtcpei 120
Qy 121 PKGSKNTEVLWEECVANSVILQNNFEFTIIDLGTSRSLPQLLRNNSVVSCKTSSSC 180
Db 121 pkgskntevlwecvansvvlqnnfeftiidlgtsrslpqlrrnnsvvscktsessc 180
Qy 181 R 181
Db 181 r 181

RESULT 3
AAW95692
ID AAW95692 standard; Protein; 153 AA.
XX
XX AAW95692;

XX 08-JUN-1999 (first entry)

XX Human endogenous retrovirus IDDMK1.2-22 env protein.

Db 61 Impavqnlwlvptvpsnrftymvsgmslrprvnylqdfsygrslkfrpkgtcpkei 120
Oy 121 PKGSKNTEVLWVEECVANSVILQNNFETIID 153
Db 121 pkgskntevlwveecvansvillqnnfetiid 153
RESULT 5
AAW95693
ID AAW95693 standard; Protein; 561 AA.
XX
AC AAW95693;
XX
DT 08-JUN-1999 (first entry)
XX
DE Human endogenous retrovirus IDDMK1.2-22 env/fs-sag protein.
XX
KW IDDM; insulin-dependent diabetes mellitus; endogenous retrovirus;
KW SAG; superantigen; provirus; autoimmune disease; type 1 diabetes;
KW diagnosis; env; envelope.
XX
OS Homo sapiens.
XX
PN EP893691-A1.
XX
PD 27-JAN-1999.
XX
PE 23-JUL-1997; 97EP-0401773.
XX
PR 23-JUL-1997; 97EP-0401773.
XX
PA (MACH)/ MACH B F.
XX
PI Conrad B, Mach B;
DR WPI; 1999-097928/09.
DR N-PSDB; AAX07515.
XX
XX Diagnosing human autoimmune disease by detecting retrovirus with
PT superantigen activity - new retrovirus associated with type 1
PT diabetes, its proviral DNA, and related vectors, transformed cells,
PT proteins, antibodies and specific binding agents, used for treating
PT or preventing autoimmune disease
XX
PS Claim 31; Fig 7E; 92pp; English.
XX
XX The sequence is that of an insulin-dependent diabetes mellitus
CC associated human endogenous retrovirus (IDDMK1.2-22) env/fs-sag protein.
CC The retrovirus has Superantigen (SAG) activity. It can be used
CC as part of a method is specifically used to diagnose type 1 diabetes
CC mellitus. Modified proteins expressed by the retroviral sequence
CC (without SAG activity but still able to induce an immune response)
CC are useful in vaccines to treat or prevent SAG-related autoimmune
CC disease; nucleic acid sequences encoding (modified) SAG can be used
CC similarly to treat such diseases. Retroviral-encoded SAG are important
CC in pathogenesis of autoimmune disease, probably by activating
CC autoreactive T cells. The method is very specific (it can differentiate
CC between expressed and non-expressed viral nucleic acids) and can be used
CC even where the pathogen is an ubiquitous endogenous retrovirus. Blood
CC or plasma samples can be tested without extensive preparation and
CC diagnosis can be made before clinical signs are apparent, allowing
CC early intervention before severe tissue damage has occurred.
XX
SQ Sequence 561 AA;
Query Match 85.9%; Score 840; DB 20; Length 561;
Best Local Similarity 100.0%; Pred. No. 4.2e-89;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 MVTPTWMDNPLEVYVNDVWVGPTDDRCAPKEEGMMINISIGYHYPPICIGRAPGC 60
|||||

Db 1 mvtptwmdnplevyvndsvwvgptddrcapkeegmmminisigyhyppicigrapgc 60
Oy 61 LMPAVQNWLVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYGRSLKFRPKGKTCPKKEI 120
|||||
Db 61 Impavqnlwlvptvpsnrftymvsgmslrprvnylqdfsygrslkfrpkgtcpkei 120
Oy 121 PKGSKNTEVLWVEECVANSVILQNNFETIID 153
Db 121 pkgskntevlwveecvansvillqnnfetiid 153
RESULT 6
AAW97746
ID AAW97746 standard; Protein; 561 AA.
XX
AC AAW97746;
XX
DT 21-MAY-1999 (first entry)
XX
DE Human endogenous retrovirus IDDMK1.2-22 envelope protein.
XX
KW HERV; IDDMK1.2-22; superantigen; SAG; antigen; IDDM;
KW insulin-dependent diabetes mellitus; autoimmune disease; diagnosis;
KW therapy; vaccine; envelope protein; env gene.
XX
OS Human endogenous retrovirus.
XX
FH Key Location/Qualifiers
ET Modified-site 17...19
FT /note= "Asn is N-glycosylated"
FT Modified-site 42...45
FT /note= "Asn is N-glycosylated"
XX
PN WO9905527-A2.
XX
PD 04-FEB-1999.
XX
PE 22-JUL-1998; 98WO-EP04926.
XX
PR 23-JUL-1997; 97EP-0401773.
XX
PR 22-JUL-1997; 97EP-0112482.
XX
PA (MEDI-) MEDIGEN SA.
XX
PI Conrad B, Mach B;
DR WPI; 1999-143118/12.
XX
XX New isolated human endogenous retrovirus - used to develop products
CC for the diagnosis, prevention and treatment of autoimmune disease,
CC particularly insulin dependent diabetes mellitus
XX
XX Claim 6; Fig 7E; 165pp; English.
CC
CC This is the envelope protein (Env) of a new human endogenous
CC retrovirus (HERV), designated IDDMK1.2-22, that has been identified
CC as the source of superantigen (SAG) activity in insulin-dependent
CC diabetes mellitus (IDDM) patients. The endogenous retrovirus is
CC ubiquitous in the human genome but is only expressed in diabetic
CC individuals. The HERV encodes SAG activity within the env gene.
CC A claimed process for the diagnosis, including the pre-symptomatic
CC diagnosis, of a human autoimmune disease associated with a HERV
CC having SAG activity comprises specifically detecting in a
CC biological sample either: (a) the mRNA of an expressed HERV having
CC SAG activity (especially IDDMK1.2-22 5'LTR, 3' long terminal
CC repeat, env or pol); (b) a protein or peptide expressed by the HERV
CC (see AAW97745-48); (c) antibodies specific to the proteins expressed
CC by the HERV; or (d) SAG activity specifically associated with the
CC HERV. Products of the invention can be used to identify substances
CC capable of blocking transcription or translation of SAG-encoding
CC nucleic acid sequences, useful in therapy and/or prevention of
CC autoimmune disease associated with the SAG. A nucleic acid encoding
CC human retroviral SAG can be used as a DNA vaccine. Expression of

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 48 AA;
XX

Query Match 22.5%; Score 220; DB 22; Length 48;
Best Local Similarity 83.3%; Pred. No. 2.7e-18;
Matches 40; Conservative 1; Mismatches 7; Indels

Db 1 Ycypciclrqapcglmpttqnlwlvveptvsatsqftyhrrvsqmslrpq 488

RESULT 8
AAM26458

AA
AC
AAM26458;
XX
XX

XX Referring #4355 encoded by probe for measuring placental
 YY
 KW Probe: microarray; human: placenta; antenatal diagnosis:

OS Homo sapiens.
XX

09-AUG-2001.
XX
XX
PD
....

PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.

PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.

PA (MOLE-) MOLECULAR DYNAMICS I
XX

DR WPI; 2001-488897/53.
XX
DM Human genome-derived clones

XX
PS
YY

CC see AA131313-AA131346). The probes are used for predicting, measuring and detecting the presence of the virus in the blood.

CC human genetic disorders.
XX
SQ Sequence 48 AA:

Query Match	22.5%
Best Local Similarity	83.3%

Qy 47 YHYPICLGRAPGCLMPAVQNWIL

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 48 AA;

Query Match 22.5%; Score 220; DB 22; Length 48;
Best Local Similarity 83.3%; Pred. No. 2.7e-18;
Matches 40; Conservative 1; Mismatches 7; Indels

Db 1 Ycypciclrqapcglmpttqnlwlvveptvsatsqftyhrvsqmslrpq 488

RESULT 8
AAM26458

AA
AC
AAM26458;
XX
XX

XX Probe: microarray; human: placenta; antenatal diagnosis:
 XX reference #435 encoded by probe for measuring placental
 KW

OS Homo sapiens.
XX

09-AUG-2001.
XX
XX
PD
....

PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.

PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.

PA (MOLE-) MOLECULAR DYNAMICS I
XX

DR WPI; 2001-488897/53.
XX
DM Human genome-derived clones

XX
PS
YY

CC see AA131313-AA131346). The probes are used for predicting, measuring and detecting the presence of the virus.

CC human genetic disorders.
XX
SQ Sequence 48 AA:

Query Match	22.5%
Best Local Similarity	83.3%

Qy 47 YHYPICLGRAPGCLMPAVQNWIL


```

XX PD 23-JUN-1994.
XX PF 07-DEC-1993; 93WO-US11907.
XX PR 07-DEC-1992; 92US-0991867.
XX PA (UYFL ) UNIV FLORIDA.
XX PI Gruidl ME, Hall RL, Moyer RW;
XX DR WPI; 1994-217887/26.
XX DR N-PSDB; AAQ66808.
XX PT New entomopoxvirus polynucleotide sequences, proteins and vectors
XX PT - are used for expression of heterologous proteins in both insect
XX PT and mammalian host cells
XX PS Disclosure; Page 68-71; 118pp; English.
XX CC The sequence of the Amsacta moorei entomopoxvirus spheroidin gene
XX CC and its flanking regions was determined. The spheroidin gene
XX CC can be used as the location for the insertion of heterologous DNA
XX CC in insect and mammalian expression systems.
XX SQ Sequence 1003 AA;

Query Match 8.2%; Score 80.5; DB 15; Length 1003;
Best Local Similarity 24.3%; Pred. No. 3.8;
Matches 27; Conservative 16; Mismatches 39; Indels 29; Gaps 5;

QY 60 CLMPAVQ-----NWLVETVTPNSRFTYHVMVG---MSLRPRVNYLQDFSYQSLKFR 110
Db 441 clkpvpknrlwlgwldcdt-----srfikhmadgsddldldvrln----- 482

QY 111 PKGKTCPE-IPGSKNTEVLVWEECVANSVILQNNFEGTII DLGTSRSI 160
Db 483 -rndiclkqalkghytnvileantypnctlsignnrfnnvfmndnktl 532

RESULT 12
AAW41301
ID AAW41301 standard; Protein; 1003 AA.
AC AAW41301;
XX DT 20-MAY-1998 (first entry)
XX DE Amsacta moorei entomopoxvirus spheroidin protein G5R.
XX KW Entomopoxvirus; spheroidin gene; AmsEPV; thymidine kinase; promoter;
XX KW insect control; viral vaccine.
XX OS Amsacta moorei.
XX PN US5721352-A.
XX PD 24-FEB-1998.
XX PF 22-NOV-1993; 93US-0107755.
XX PR 22-NOV-1993; 93US-0107755.
XX PR 19-FEB-1991; 91US-0657584.
XX PR 30-JAN-1992; 92US-0827685.
XX PR 12-FEB-1992; 92WO-US00855.
XX PA (UYFL ) UNIV FLORIDA RES FOUND.
XX PI Gruidl ME, Hall RL, Moyer RW;
XX PS WPI; 1998-168476/15.
XX DR N-PSDB; AAV14507, AAV14517.

XX PD 23-JUN-1994.
XX PF 07-DEC-1993; 93WO-US11907.
XX PR 07-DEC-1992; 92US-0991867.
XX PS Disclosure; Columns 45-50; 55pp; English.
XX CC This sequence is encoded by the Amsacta moorei entomopoxvirus (AmsEPV)
XX CC spheroidin gene, which is an example of the gene of the
XX CC invention, which encodes a 115 kDa protein. EPV spheroidin and
XX CC thymidine kinase promoters can be used in DNA constructs and vectors for
XX CC expression of heterologous genes in insects or mammalian cells,
XX CC e.g. vectors containing Bacillus thuringiensis toxin genes for use in
XX CC insect control, or recombinant vaccinia or swinepox viruses for use as
XX CC viral vaccines.
XX SQ Sequence 1003 AA;

Query Match 8.2%; Score 80.5; DB 19; Length 1003;
Best Local Similarity 24.3%; Pred. No. 3.8;
Matches 27; Conservative 16; Mismatches 39; Indels 29; Gaps 5;

QY 60 CLMPAVQ-----NWLVETVTPNSRFTYHVMVG---MSLRPRVNYLQDFSYQSLKFR 110
Db 441 clkpvpknrlwlgwldcdt-----srfikhmadgsddldldvrln----- 482

QY 111 PKGKTCPE-IPGSKNTEVLVWEECVANSVILQNNFEGTII DLGTSRSI 160
Db 483 -rndiclkqalkghytnvileantypnctlsignnrfnnvfmndnktl 532

RESULT 13
AAY30169
ID AAY30169 standard; Protein; 1003 AA.
XX AC AAY30169;
XX DT 28-OCT-1999 (first entry)
XX DE Spheroidin protein encoded by Amsacta moorei entomopoxvirus genome.
XX KW Spheroidin; Entomopoxvirus; expression system; replication;
XX KW heterologous gene expression; thymidine kinase; poxvirus; vaccinia;
XX KW swinepox virus; insect pest control; immunity.
XX OS Amsacta moorei entomopoxvirus.
XX PN US5935777-A.
XX PD 10-AUG-1999.
XX PF 17-OCT-1995; 95US-0544332.
XX PR 17-OCT-1995; 95US-0544332.
XX PR 19-FEB-1991; 91US-0657584.
XX PR 30-JAN-1992; 92US-0827685.
XX PR 12-FEB-1992; 92WO-US00855.
XX PR 07-DEC-1992; 92US-0991867.
XX PA (UYFL ) UNIV FLORIDA RES FOUND INC.
XX PI Gruidl ME, Hall RL, Li Y, Moyer RW;
XX DR WPI; 1999-457596/38.
XX DR N-PSDB; AA210081.
XX PT Novel expression system for the expression of heterologous sequences
XX PT in insect and mammalian host cells
XX PS Disclosure; Column 61-68; 72pp; English.
XX CC AAY30165-70 represent proteins encoded by open reading frames (ORFs)

```


PT causes von Willebrand's disease in canines -

XX

5d

XX

The present sequence is a canine prep

The present sequence is a canine prepro-von Willebrand factor (vWF) having homology with the human prepro-vWF sequence. Mutation in the vWF gene causes von Willebrand's disease (vWD) in dogs. The present sequence is used in diagnostic tests for detecting carriers of the mutation that causes von Willebrand's disease in canines such as Scottish terriers, Doberman pinschers, Shetland sheepdogs, Manchester terriers and Poodles. Such tests may be used by breeders to reduce the frequency of the disease-causing allele and the incidence of disease. The vWF cDNA sequences may be used to determine the genetic defect that causes vWD in other breeds as well as other species.

XX

SQ Sequence 2813 AA;

Query Match	8.2%;	Score	80.5;	DB	21;	Length	2813;
Best Local Similarity	21.8%;	Pred. No.	16;				
Matches	46;	Conservative	29;	Mismatches	73;	Indels	63;
Gaps	8;						
Qy	26	TDDRCPAKPEEGMINIS---	IGYHYPPIC-----	LGRAPCCLMPAONWLVE	VPT	74	
Db	859	tdhvcdatcsaigmahyltfdgkylfpgceqyvlvqdy	csgnptirilvgnecg	yps	918		
Qy	75	VSPNSRTYTHMVSGLRPRVNYLQDSYQBSLSKRP	KPGKTCPEIKPGSKNTEV	L----	130		
Db	919	vkckkrvti-lveg-----	geielfdgevnvkkpmkdethfev	esggqyvilllgkal	970		
Qy	131	--VNEECVANSVWT-----	LNNEGTI-----	IDLGTSSRLPQ	163		
Db	971	svvwdhrlsivtkrtygeqvcqlcnfdg	iqndftssslqieedp	vdfignkwkvpq	1030		
Qy	164	LLRT-----	NSVVSCKTSSSCR	181			
Db	1031	cadtkkvpldsspavchhnmkqtmvdsscr	1061				

Search completed: April 9, 2002, 16:56:45

Job time: 479 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:06:33 ; Search time 70.84 Seconds
(without alignments)

Title: US-09-490-700-40
Perfect score: 978
Sequence: 1 MVTPVTWMDNPVEV

Scoring table: BLOSUM62

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```

SPREMBL_17:*
1: sp_bacteria:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %		Length	DB	ID	Description
	Match	Score				
1	840	85.9	153	4	Q95280	Q95280 homo sapien
2	840	85.9	153	12	Q42043	Q42043 unclassified
3	831	85.0	560	12	Q9QC06	Q9QC06 human endog
4	811	82.9	153	4	Q9UNW3	Q9UNW3 homo sapien
5	805	82.3	153	4	Q9UNW2	Q9UNW2 homo sapien
6	805	82.3	245	4	Q9HDB8	Q9HDB8 homo sapien
7	792	81.0	1361	4	Q14273	Q14273 homo sapien
8	789	80.7	2294	4	Q9UKH9	Q9UKH9 homo sapien
9	784	80.2	153	4	Q95284	Q95284 homo sapien
10	782	80.0	153	4	Q95281	Q95281 homo sapien
11	779	79.7	694	12	Q9YNA9	Q9YNA9 human endog
12	779	79.7	699	4	Q9UBU4	Q9UBU4 homo sapien
13	779	79.7	699	12	Q69384	Q69384 human endog
14	778	79.6	514	12	Q69386	Q69386 human endog
15	778	79.6	694	12	Q9YNA7	Q9YNA7 human endog
16	777	79.4	153	4	Q95282	Q95282 homo sapien
17	777	79.4	694	12	Q9YNA5	Q9YNA5 human endog
18	774	79.1	277	4	Q9UKH7	Q9UKH7 homo sapien
19	752.5	76.9	698	4	Q9UKH3	Q9UKH3 homo sapien

20	751.5	76.8	152	4	O95283	O95283	homo sapien
21	621	50.2	124	4	O9HDC0	O9HDC0	homo sapien
22	491	50.2	206	12	O71037	O71037	human endog
23	472	48.3	206	12	O71072	O71072	human endog
24	117.5	12.0	300	6	O9GMP3	O9GMP3	macaca fasc
25	87.5	8.9	1785	12	O9DU47	O9DU47	chiba virus
26	83.5	8.5	990	10	O80770	O80770	arabidopsis
27	83	8.5	347	11	O08521	O08521	cricetulus
28	83	8.5	843	6	O9RTQ3	O9RTQ3	equus scrofa
29	82	8.4	226	7	O30489	O30489	equus caball
30	81	8.3	633	5	O9VT54	O9VT54	drosophila
31	81	8.3	972	2	O9RL94	O9RL94	staphylococ
32	81	8.3	973	2	O48348	O48348	enterococcu
33	79	8.1	295	10	O22158	O22158	arabidopsis
34	79	8.1	492	5	O9XX15	O9XX15	caenorhabdi
35	79	8.1	514	5	O9XX16	O9XX16	caenorhabdi
36	77.5	7.9	478	10	O9FG60	O9FG60	arabidopsis
37	76.5	7.8	711	10	O9SPM1	O9SPM1	lycopersico
38	76	7.8	660	11	O9RLA8	O9RLA8	mus musculus
39	76	7.8	704	3	O9P7Y8	O9P7Y8	schizosacch
40	75.5	7.7	318	2	O55633	O55633	synecocyst
41	75	7.7	971	2	O56407	O56407	listeria mo
42	74.5	7.6	623	10	O9C9B4	O9C9B4	arabidopsis
43	74	7.6	459	12	P89307	P89307	yam mosaic
44	74	7.6	3885	5	O9N533	O9N533	caenorhabdi
45	73.5	7.5	297	2	O9XA09	O9XA09	streptomyce

ALIGNMENTS

RESULT	1
O95280	
ID	O95280 PRELIMINARY; PRT; 153 AA.
AC	O95280;
DT	01-MAY-1999 (TrEMBLrel. 10, Created)
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT	01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE	ENVELOPE PROTEIN RIC-1.
GN	ENV.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
[1]	
RN	SEQUENCE FROM N.A.
RP	MEDLINE=98449692; PubMed=9778243;
RX	Lan M.S., Mason A., Coutant R., Chen Q.Y., Vargas A., Rao J.,
RA	Gomez R., Chalew S., Garry R., MacIaren N.K.:
RA	"HERV-K10s and immune-mediated (type 1) diabetes.";
RT	Cell 95:14-16(1998).
RL	[2]
RN	SEQUENCE FROM N.A.
RP	Hasuike S., Jinno Y.;
RT	"Isolation and localization of an endogenous retrovirus gene, a
RT	candidate gene for type 1 diabetes, and identification of a CA repeat
RT	marker at its locus";
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF084864; AAC68893.1; -
DK	EBML; AF134984; AAC63055.1; -
KW	Envelope protein.
SQ	SEQUENCE 153 AA; 173305 MW; 2B36E1656DF0C01A CRC64;

	Query Match	85.9%	Score 840;	DB 4;	Length 153;
	Best Local Similarity	100.0%	Pred. No. 5.6e-79;		
	Matches 153;	Conservative	0;	Mismatches	0;
Qy	1	MVPTVTWMDNP	IEYVYVND	SVVMPGTTD	DRCPAKPEEGMMINISIGYHYPPICLGRAPGC
Db	1	MVPTVTWMDNP	IEYVYVND	SVVMPGTTD	DRCPAKPEEGMMINISIGYHYPPICLGRAPGC
Qy	61	LMFAOVNWLVE	VPVTSPNSRFTYH	VMVSGMSLR	PRVNYLDDFSYORSLSKLFKPKGTCKPEI

```
|||||
Db 61 LMPAVQNNLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120
Qy 121 PGSKNTEVLWEECVANSVILQNNFEFTIID 153
Db 121 PGSKNTEVLWEECVANSVILQNNFEFTIID 153

RESULT 2
O42043 PRELIMINARY; PRT; 153 AA.
AC O42043;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE ENVELOPE PROTEIN.
GN ENV.
OS unclassified Retroviridae.
OC Viruses; Retroviral viruses; Retroviridae.
OX NCBI_TaxID=35276;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97386455; PubMed=9244304;
RA Conrad B., Weissman R.N., Boni J., Arcari R., Schubach J., Mach B.;
RT "A human endogenous retroviral superantigen as candidate autoimmune
RL gene in type I diabetes.";
RL Cell 90:303-313(1997).
DR EMBL; AF012337; AAC58457.1; -.
DR EMBL; AF012336; AAC58456.1; -.
KW Envelope protein.
SQ SEQUENCE 153 AA; 17305 MW; 2B36E1656DF0C01A CRC64;

Query Match 85.9%; Score 840; DB 12; Length 153;
Best Local Similarity 100.0%; Pred. No. 5.6e-79;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVTPTWMDNPIEVYVNDVSVWVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60
Db 1 MVTPTWMDNPIEVYVNDVSVWVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60

Qy 61 LMPAVQNNLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120
Db 61 LMPAVQNNLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120

Qy 121 PGSKNTEVLWEECVANSVILQNNFEFTIID 153
Db 121 PGSKNTEVLWEECVANSVILQNNFEFTIID 153

RESULT 3
O9QC06 PRELIMINARY; PRT; 560 AA.
AC O9QC06;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ENV PROTEIN.
GN ENV.
OS Human endogenous retrovirus K.
OC Viruses; Retroviral viruses; Retroviridae.
OX NCBI_TaxID=45617;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99445825; PubMed=10516026;
RA Toenjes R.R., Czauderna F., Kurth R.;
RT "Genome wide screening, cloning, chromosomal assignment and expression
RL of full-length human endogenous retrovirus type K (HERV-K).";
RL J. Virol. 73:9187-9195(1999).
DR EMBL; Y18890; CAB56604.1; -.
SQ SEQUENCE 560 AA; 63671 MW; 8A4565663901BC3A CRC64;
```

```
Query Match 85.0%; Score 831; DB 12; Length 560;
Best Local Similarity 99.3%; Pred. No. 2.2e-77;
Matches 152; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MVTPTWMDNPIEVYVNDVSVWVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60
Db 1 MVTPTWMDNPIEVYVNDVSVWVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60

Qy 61 LMPAVQNNLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120
Db 61 LMPAVQNNLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120

Qy 121 PGSKNTEVLWEECVANSVILQNNFEFTIID 153
Db 121 PGSKNTEVLWEECVANSVILQNNFEFTIID 153

RESULT 4
O9UNW3 PRELIMINARY; PRT; 153 AA.
AC O9UNW3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE ENVELOPE PROTEIN RIC-2.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449692; PubMed=9778243;
RA Lan M.S., Mason A., Coutant R., Chen Q.Y., Vargas A., Rao J.,
RA Gomez R., Chalew S., Garry R., Maclaren N.K.;
RT "HERV-K10s and immune-mediated (type 1) diabetes.";
RL Cell 95:14-16(1998).
DR EMBL; AF084865; AAC68894.1; -.
KW Envelope protein.
SQ SEQUENCE 153 AA; 17226 MW; 4632F483BFD9517A CRC64;

Query Match 82.9%; Score 811; DB 4; Length 153;
Best Local Similarity 96.7%; Pred. No. 5.5e-76;
Matches 146; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MVTPTWMDNPIEVYVNDVSVWVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60
Db 1 MVTPTWMDNPIEVYVNDVSVWVPGPTDDRCPCAKPEEGMMINISIGYHYPPICLGRAPGC 60

Qy 61 LMPAVQNNLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120
Db 61 LMPAVQNNLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120

Qy 121 PGSKNTEVLWEECVANSVILQNNFEFTIID 153
Db 121 PGSKNTEVLWEECVANSVILQNNFEFTIID 153

RESULT 5
O9UNW2 PRELIMINARY; PRT; 153 AA.
AC O9UNW2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE ENVELOPE PROTEIN RIC-7.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
```


RP SEQUENCE FROM N.A.
 RX MEDLINE-98449692; PubMed-9778243;
 RA Lan M.S., Mason A., Coutant R., Chen Q.Y., Vargas A., Rao J.,
 RA Gomez R., Chalew S., Garry R., MacLaren N.K.:
 RT "HERV-K10s and immune-mediated (type 1) diabetes.";
 RL Cell 95:14-16(1998).
 DR EMBL: AF084870; AAC68899.1; -.
 KW Envelope protein.
 SQ SEQUENCE 153 AA; 17406 MW; 90B91BCE0040C5A3 CRC64;

Query Match 82.3%; Score 805; DB 4; Length 153;
 Best Local Similarity 96.7%; Pred. No. 2.3e-75;
 Matches 148; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MVTPTVMDNPIEVYVNDVSVVWPGPTDRCAPKEEGMMINISIGVHYPPICLGRAPGC 60
 DB 1 MVTPTVMDNPIEVYVNDVSVVWPGPTDRCAPKEEGMMINISIVYRPPICLGRAPGC 60

QY 61 LMPAVQNLVLEVPVTPNSRFTYHVMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120
 DB 61 LMPAVQNLVLEVPVTPNSRFTYHVMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120

QY 121 PKGSKNTEVLWEECVANSVVIQNNFEFTIID 153
 DB 121 PKGSKNTEVLWEECVANSVVIQNNFEFTIID 153

RESULT 6
 Q9HDB8 PRELIMINARY; PRT; 245 AA.

AC Q9HDB8
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE ENV.
 GN ENV.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RA Jinno Y., Sugimoto J.;
 RP "Human endogenous retrovirus HERV-K(II) and flanking sequences.";
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB047240; BAB11760.1; -.
 SQ SEQUENCE 245 AA; 27904 MW; 198F26D65ED56DDB CRC64;

Query Match 82.3%; Score 805; DB 4; Length 245;
 Best Local Similarity 96.7%; Pred. No. 4e-75;
 Matches 148; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MVTPTVMDNPIEVYVNDVSVVWPGPTDRCAPKEEGMMINISIGVHYPPICLGRAPGC 60
 DB 1 MVTPTVMDNPIEVYVNDVSVVWPGPTDRCAPKEEGMMINISIVYRPPICLGRAPGC 60

QY 61 LMPAVQNLVLEVPVTPNSRFTYHVMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120
 DB 61 LMPAVQNLVLEVPVTPNSRFTYHVMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120

QY 121 PKGSKNTEVLWEECVANSVVIQNNFEFTIID 153
 DB 121 PKGSKNTEVLWEECVANSVVIQNNFEFTIID 153

RESULT 7
 Q14273 PRELIMINARY; PRT; 1361 AA.

AC Q14273
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE POL/ENV ORF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=LIVER;
 RX MEDLINE-87036922; PubMed-3021993;
 RA Ono M., Yasunaga T., Miyata T., Ushikubo H.;
 RT "Nucleotide sequence of human endogenous retrovirus genome related to
 RT the mouse mammary tumor virus genome.";
 RL J. Virol. 60:589-598(1986).
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
 CC TRANSCRIPTASE).
 DR EMBL: M14123; AAA88033.1; -.
 DR InterPro: IPR001037; Integrase_C.
 DR InterPro: IPR003308; Integrase_Zn.
 DR InterPro: IPR002156; RNaseH.
 DR InterPro: IPR001584; Rve.
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF00552; Integrase; 1.
 DR Pfam: PF02022; Integrase_Zn; 1.
 DR Pfam: PF00075; RNaseH; 1.
 DR Pfam: PF00665; Rve; 1.
 DR Pfam: PF00078; rvt; 1.
 KW RNA-directed DNA polymerase.
 SQ SEQUENCE 1361 AA; 153797 MW; CEB91B3F407B9498 CRC64;

Query Match 81.0%; Score 792; DB 4; Length 1361;
 Best Local Similarity 94.8%; Pred. No. 6.5e-73;
 Matches 145; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 MVTPTVMDNPIEVYVNDVSVVWPGPTDRCAPKEEGMMINISIGVHYPPICLGRAPGC 60
 DB 774 MVTPTVMDNPIEVYVNDVSVVWPGPTDRCAPKEEGMMINISIGVHYPPICLGRAPGC 833

QY 61 LMPAVQNLVLEVPVTPNSRFTYHVMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120
 DB 834 LMPAVQNLVLEVPVTPNSRFTYHVMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 893

QY 121 PKGSKNTEVLWEECVANSVVIQNNFEFTIID 153
 DB 894 PKGSKNTEVLWEECVANSVVIQNNFEFTIID 926

RESULT 8
 Q9UKH9 PRELIMINARY; PRT; 2294 AA.

ID Q9UKH9
 AC Q9UKH9
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE GAG-PRO-POL-ENV PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-99400989; PubMed-10469592;
 RA Barbulescu M., Turner G., Seaman M.I., Deinard A.S., Kidd K.K.,
 RA Lenz J.;
 RT "Many human endogenous retrovirus K (HERV-K) proviruses are unique to
 RT humans.";
 RL Curr. Biol. 9:861-868(1999).
 CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
 CC TRANSCRIPTASE).
 CC -1- SIMILARITY: BELONGS TO ZN-FINGER CCHC TYPE FAMILY.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO THE PEPTIDASE FAMILY A1; ALSO
 CC KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.

```
DR EMBL; AF164611; AAD51793.1; -.
DR HSSP; P16088; 1DUT.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001995; Asp_prot_retrov.
DR InterPro; IPR001428; DUTPase.
DR InterPro; IPR003322; Gag_p10.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR000467; G_patch.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_zn.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; Rvtse.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00692; dUTPase; 1.
DR Pfam; PF01585; G_patch; 1.
DR Pfam; PF02337; Gag_p10; 1.
DR Pfam; PF00607; gag_p24; 1.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF02022; Integrase_zn; 1.
DR Pfam; PF00075; rnaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; ivp; 1.
DR Pfam; PF00078; ivt; 1.
DR Pfam; PF00098; zf_CCHC; 2.
DR ProDom; PD000946; dUTPase; 1.
DR SMART; SM00443; G_patch; 1.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS00175; ASP_PROT_RETROV; 1.
DR Aspartyl protease; Hydrolase; RNA-directed DNA polymerase;
KW Zinc-finger.
SQ SEQUENCE 2294 AA; 257099 MW; 425C81FDDF8A80D0 CRC64;

Query Match 80.7%; Score 789; DB 4; Length 2294;
Best Local Similarity 94.8%; Pred. No. 2.5e-72;
Matches 145; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 MVTPTWMDNPIEVYVNDSDVWVPGPTDDRCAPKPEEGGMINISIGYHYPPICLGRAPGC 60
Db 1795 MVTPTWMDNPIEVYVNDSDVWVPGPTDDRCAPKPEEGGMINISIGYHYPPICLGRAPGC 1854

Qy 61 LMPAVQNLVLEVPVSPNSRFTYHMVSGMSLRPRVNYLQDFSQYRSKLFKRPKGTCPKEI 120
Db 1855 LMPAVQNLVLEVPVSPISRTYHMVSGMSLRPRVNYLQDFSQYRSKLFKRPKGTCPKEI 1914

Qy 121 PGSKNTEVLVWEECVANSVILQNNFEGTIID 153
Db 1915 PGSKNTEVLVWEECVANSVILQNNFEGTIID 1947

RESULT 9
Q95284
ID Q95284 PRELIMINARY; PRT; 153 AA.
AC Q95284;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ENVELOPE PROTEIN RIC-6.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449692; PubMed=9778243;
RA Lan M.S., Mason A., Coutant R., Chen Q.Y., Vargas A., Rao J.,
RA Gomez R., Chalew S., Garry R., Maclaren N.K.;
RT "HERV-K10s and immune-mediated (type 1) diabetes.";
RL Cell 95:14-16(1998).
KW EMBL; AF084869; AAC68898.1; -.

KW Envelope protein.
SQ SEQUENCE 153 AA; 17391 MW; 4088880E06FB7677 CRC64;

Query Match 80.2%; Score 784; DB 4; Length 153;
Best Local Similarity 94.1%; Pred. No. 3.4e-73;
Matches 144; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 MVTPTWMDNPIEVYVNDSDVWVPGPTDDRCAPKPEEGGMINISIGYHYPPICLGRAPGC 60
Db 1 MVTPTWMDNPIEVYVNDSDVWVPGPTDDRCAPKPEEGGMINISIGYHYPPICLGRAPGC 60

Qy 61 LMPAVQNLVLEVPVSPNSRFTYHMVSGMSLRPRVNYLQDFSQYRSKLFKRPKGTCPKEI 120
Db 61 LMPAVQNLVLEVPVSPISRTYHMVSGMSLRPRVNYLQDFSQYRSKLFKRPKGTCPKEI 120

Qy 121 PGSKNTEVLVWEECVANSVILQNNFEGTIID 153
Db 121 PGSKNTEVLVWEECVANSVILQNNFEGTIID 153

RESULT 10
Q95281
ID Q95281 PRELIMINARY; PRT; 153 AA.
AC Q95281;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ENVELOPE PROTEIN RIC-3.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449692; PubMed=9778243;
RA Lan M.S., Mason A., Coutant R., Chen Q.Y., Vargas A., Rao J.,
RA Gomez R., Chalew S., Garry R., Maclaren N.K.;
RT "HERV-K10s and immune-mediated (type 1) diabetes.";
RL Cell 95:14-16(1998).
KW EMBL; AF084866; AAC68895.1; -.

KW Envelope protein.
SQ SEQUENCE 153 AA; 17360 MW; 28C01AD348C6D806 CRC64;

Query Match 80.0%; Score 782; DB 4; Length 153;
Best Local Similarity 94.1%; Pred. No. 5.4e-73;
Matches 144; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MVTPTWMDNPIEVYVNDSDVWVPGPTDDRCAPKPEEGGMINISIGYHYPPICLGRAPGC 60
Db 1 MVTPTWMDNPIEVYVNDSEWVPGPTDDRCAPKPEEGGMINISIGYHYPPICLGTAPGC 60

Qy 61 LMPAVQNLVLEVPVSPNSRFTYHMVSGMSLRPRVNYLQDFSQYRSKLFKRPKGTCPKEI 120
Db 61 LMPAVQNLVLEVPVSPISRTYHMVSGMSLRPRVNYLQDFSQYRSKLFKRPKGTCPKEI 120

Qy 121 PGSKNTEVLVWEECVANSVILQNNFEGTIID 153
Db 121 PGSKNTEVLVWEECVANSVILQNNFEGTIID 153

RESULT 11
Q9YNA9
ID Q9YNA9 PRELIMINARY; PRT; 694 AA.
AC Q9YNA9;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ENV PROTEIN.
GN ENV.
OS Human endogenous retrovirus K.
```

OC Viruses; Retroid viruses; Retroviridae.
 OX NCBI_TaxID=45617;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Toenjes R.R., Czauderna F., Kurth R.;
 RT "Full-length human endogenous retrovirus type K (HERV-K) elements
 RT encoding Gag, Pol and Env proteins are localised on chromosomes 7 and
 RT 19.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Y17832; CAA76880.1; -
 SQ SEQUENCE 694 AA; 78690 MW; 82F91825669CF25B CRC64;

Query Match 79.7%; Score 779; DB 12; Length 694;
 Best Local Similarity 92.8%; Pred. No. 6.5e-72;
 Matches 142; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 MVTPTVWMDNPVYVNDVSVVWVPGPTDDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 60
 Db 107 LIRAVTWMNDNPTEYVYVNDVSVVWVPGPTDDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 166
 QY 61 LMPAVQNLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPCKEI 120
 Db 167 LMPAVQNLVVEPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPCKEI 226
 QY 121 PGSKNTEVLWEECVANSVILQNNFEFTIID 153
 Db 227 PKSKNTEVLWEECVANSVILQNNFEFTIID 259

Query Match 79.7%; Score 779; DB 12; Length 694;
 Best Local Similarity 92.8%; Pred. No. 6.5e-72;
 Matches 142; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 MVTPTVWMDNPVYVNDVSVVWVPGPTDDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 60
 Db 107 LIRAVTWMNDNPTEYVYVNDVSVVWVPGPTDDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 166
 QY 61 LMPAVQNLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPCKEI 120
 Db 167 LMPAVQNLVVEPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPCKEI 226
 QY 121 PGSKNTEVLWEECVANSVILQNNFEFTIID 153
 Db 227 PKSKNTEVLWEECVANSVILQNNFEFTIID 259

RESULT 12
 Q9UBU4
 ID Q9UBU4 PRELIMINARY; PRT; 699 AA.
 AC Q9UBU4;
 DT 01-MAY-2000 (TEMBLrel. 13, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
 DE ENVELOPE PROTEIN.
 GN ENV.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99400989; PubMed=10469592;
 RA Barbulescu M., Turner G., Seaman M.I., Delnard A.S., Kidd K.K.,
 RA Lenz J.;
 RT "Many human endogenous retrovirus K (HERV-K) proviruses are unique to
 RL humans.";
 RL Curr. Biol. 9:861-868(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99178257; PubMed=10080172;
 RA Mayer J., Sauter M., Racz A., Scherer D., Mueller-Lantsch N.,
 RA Meese E.;
 RT "An almost-intact human endogenous retrovirus K on human chromosome
 RL 7.";
 RL Nat. Genet. 21:257-258(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Reus K., Mayer J., Sauter M., Scherer D., Mueller-Lantsch N.,
 RA Meese E.;
 RT "Further characterization of the almost intact human endogenous
 RL retrovirus K on human chromosome 7.";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF164614; AAD51798.1; -
 DR EMBL: AF074086; AAF88168.1; -
 DR EMBL: AF074086; AAD21098.1; -
 KW Envelope protein.
 SQ SEQUENCE 699 AA; 79217 MW; 3844DA6D80D76DFC CRC64;

Query Match 79.7%; Score 779; DB 4; Length 699;
 Best Local Similarity 92.8%; Pred. No. 6.6e-72;
 Matches 142; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 MVTPTVWMDNPVYVNDVSVVWVPGPTDDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 60
 Db 112 LIRAVTWMNDNPTEYVYVNDVSVVWVPGPTDDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 171
 QY 61 LMPAVQNLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPCKEI 120
 Db 172 LMPAVQNLVVEPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPCKEI 231
 QY 121 PGSKNTEVLWEECVANSVILQNNFEFTIID 153
 Db 232 PKSKNTEVLWEECVANSVILQNNFEFTIID 264

RESULT 13
 Q69384
 ID Q69384 PRELIMINARY; PRT; 699 AA.
 AC Q69384;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
 DE ENV MRNA.
 GN ENV.
 OS Human endogenous retrovirus.
 OC Viruses; Retroid viruses; Retroviridae.
 OX NCBI_TaxID=11827;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95074858; PubMed=7983704;
 RA Loewer R., Toenjes R.R., Korbacher C., Kurth R., Loewer J.;
 RT "Identification of a Rev-related protein by analysis of spliced
 RT transcripts of the human endogenous retroviruses HTDV/HERV-K.";
 DR J. Virol. 69:141-149(1995).
 DR EMBL: X82272; CAA57723.1; -
 SQ SEQUENCE 699 AA; 79217 MW; 3844DA6D80D76DFC CRC64;

Query Match 79.7%; Score 779; DB 12; Length 699;
 Best Local Similarity 92.8%; Pred. No. 6.6e-72;
 Matches 142; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 MVTPTVWMDNPVYVNDVSVVWVPGPTDDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 60
 Db 112 LIRAVTWMNDNPTEYVYVNDVSVVWVPGPTDDRCAPKPEEGMMINISIGYHYPPICLGRAPGC 171
 QY 61 LMPAVQNLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPCKEI 120
 Db 172 LMPAVQNLVVEPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYORSLSKFRPKGKTCPCKEI 231
 QY 121 PGSKNTEVLWEECVANSVILQNNFEFTIID 153
 Db 232 PKSKNTEVLWEECVANSVILQNNFEFTIID 264

RESULT 14
 Q69386
 ID Q69386 PRELIMINARY; PRT; 514 AA.
 AC Q69386;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
 DE POL/ENV GENE (FRAGMENT).
 GN POL/ENV.
 OS Human endogenous retrovirus K.
 OC Viruses; Retroid viruses; Retroviridae.
 OX NCBI_TaxID=45617;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97213939; PubMed=9060628;
 RA Toenjes R.R., Limbach C., Lower R., Kurth R.;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:07:04 ; Search time 23.68 Seconds
(without alignments)
280.251 Million cell updates/sec

Title: US-09-490-700-40

Perfect score: 978

Sequence: 1 MVTPTWMDNPIEVVNDV.....PQLRTNSVSKTSESSCR 181

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	768.5	78.6	584	1 ENVL_HUMAN	P10267 homo sapien
2	91.5	9.4	529	1 RA12_MOUSE	Q9qy8 mus musculus
3	83	8.5	348	1 RPS3_SCHPO	O94616 schizosacch
4	82.5	8.4	530	1 RA12_HUMAN	Q9y5p3 homo sapien
5	82.5	8.4	1078	1 S24A_HUMAN	O95486 homo sapien
6	80.5	8.2	1002	1 SPHR_AMEPV	P29815 amsacta moo
7	76.5	7.8	2813	1 VWF_CANFA	Q28295 canis famill
8	76	7.8	896	1 CYRB_MOUSE	P26955 mus musculus
9	75	7.7	481	1 GLC3_SOYBN	P11828 glycine max
10	74	7.6	466	1 HN3A_RAT	P23512 rattus norv
11	73.5	7.5	598	1 P2CD_MOUSE	O9q267 mus musculus
12	73.5	7.5	1420	1 SRB9_YEAST	P38931 saccharomyc
13	72.5	7.4	602	1 PGH1_RAT	Q63921 rattus norv
14	72.5	7.4	1475	1 APU_THETY	P16950 t amylopull
15	72	7.4	688	1 ENV_MMTVB	P10259 mouse mamma
16	71.5	7.3	1222	1 PM5P_HUMAN	Q15155 homo sapien
17	71.5	7.3	1481	1 APU_THET	P38939 t amylopull
18	71	7.3	335	1 SCN1_SCHPO	P41890 schizosacch
19	71	7.3	884	1 IMB_DROME	O18388 drosophila
20	70.5	7.2	470	1 VL2_HPV54	Q81023 human papill
21	70.5	7.2	533	1 HHM1_DUGTI	Q00400 dugesia tig
22	69.5	7.1	497	1 TRPE_ACICA	P23315 acinetobact
23	69.5	7.1	555	1 MASY_PICAN	P21360 pichia anqu
24	68.5	7.0	549	1 TREF_ECOLI	P37196 escherichia
25	68	7.0	214	1 COMA_BACSU	P41204 bacillus su
26	68	7.0	575	1 NODU_RHITR	Q53515 rhizobium c
27	68	7.0	688	1 ENV_MMTVG	P03374 mouse mamma
28	68	7.0	1210	1 RPB2_SCHPO	Q02061 schizosacch
29	67.5	6.9	445	1 CTB2_MOUSE	P56546 mus musculus
30	67.5	6.9	473	1 HN3A_HUMAN	P55317 homo sapien
31	67.5	6.9	494	1 VGA_BPAL3	P25243 bacterioph
32	67.5	6.9	494	1 VGA_BPPHK	P25244 bacterioph
33	67.5	6.9	655	1 ACDV_HUMAN	P49748 homo sapien

34	67.5	6.9	1595	1 SOS_DROME	P26675 drosophila
35	67.5	6.9	1742	1 MY5C_HUMAN	Q9nqx4 homo sapien
36	67	6.9	602	1 PGH1_MOUSE	P22437 mus musculus
37	67	6.9	624	1 MTN4_MOUSE	O89029 mus musculus
38	67	6.9	840	1 YIF6_YEAST	P40322 saccharomyc
39	66.5	6.8	292	1 YG29_YEAST	P53242 saccharomyc
40	66.5	6.8	807	1 FSPQ_RAT	P35446 rattus norv
41	66.5	6.8	963	1 YQ36_CAEEL	Q09457 caenorhabdi
42	66.5	6.8	1091	1 MSH3_MOUSE	P13705 mus musculus
43	66.5	6.8	3396	1 PGCY_HUMAN	P13611 homo sapien
44	66	6.7	466	1 NMT_SCHPO	O43010 schizosacch
45	66	6.7	486	1 PRL1_ARATH	Q42384 arabidopsis

ALIGNMENTS

RESULT	1
ENVL_HUMAN	
ID	ENVL_HUMAN
AC	P10267
DT	01-MAR-1989 (Rel. 10, Created)
DT	01-MAR-1989 (Rel. 10, Last sequence update)
DT	01-MAR-1989 (Rel. 10, Last annotation update)
DE	RETROVIRUS-RELATED ENV POLYPROTEIN.
GN	ENV.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=87036922; PubMed=3021993;
RA	Ono M., Yasunaga T., Miyata T., Ushikubo H.;
RT	"Nucleotide sequence of human endogenous retrovirus genome related to
RL	the mouse mammary tumor virus genome.";
RL	J. Virol. 60:589-598(1986).
DR	PIR; E24483; VCHUER.
KW	Glycoprotein; Coat protein; Polyprotein.
FT	CARBOHYD 14 14 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 346 346 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 439 439 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 451 451 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE 584 AA; 66076 MW; FD64418F1C619166 CRC64;

Query Match 78.6%; Score 768.5; DB 1; Length 584;
Best Local Similarity 94.7%; Pred. No. 1.4e-68;
Matches 142; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY	4	PVTWMDNPIEVVNDVSVVPGPTDDRCPCAPKEEGMMINISIGYHYPPICLGRAPGCLMP	63
DB	1	PVTWMDNPIEVVNDVSVVPGPTDDRCPCAPKEEGMMINISIGYHYPPICLGRAPGCLMP	60
QY	64	AVQNLVVEPTVSPNSRFTYHYVSGMSLRPNVYQDSYQSRSLKFRPKGKTCPCKEIPKG	123
DB	61	AVQNLVVEPTVSPISRTYHYVSGMSLRPNVYQDSYQSRSLKFRPKGKTCPCKEIPKE	120
QY	124	SKNTEVLWEECVANSVVIQNNFETIID	153
DB	121	SKNTEVLWEECVANSVIL-NNEFETIID	149

RESULT	2
RA12_MOUSE	
ID	RA12_MOUSE
AC	Q9QVY8;
STANDARD;	PRT;
529	AA.

FT	DISULFID	898	993	BY SIMILARITY.	
FT	DISULFID	914	921	BY SIMILARITY.	
FT	DISULFID	1060	1084	BY SIMILARITY.	
FT	DISULFID	1071	1111	BY SIMILARITY.	
FT	DISULFID	1089	1091	BY SIMILARITY.	
FT	DISULFID	1153	1165	BY SIMILARITY.	
FT	DISULFID	1149	1169	BY SIMILARITY.	
FT	DISULFID	1126	1130	BY SIMILARITY.	
FT	DISULFID	1196	1199	BY SIMILARITY.	
FT	DISULFID	1234	1237	BY SIMILARITY.	
FT	DISULFID	1272	1458	BY SIMILARITY.	
FT	DISULFID	1669	1670	BY SIMILARITY.	
FT	DISULFID	1686	1872	BY SIMILARITY.	
FT	DISULFID	1879	1904	BY SIMILARITY.	
FT	DISULFID	1899	1940	OR 1942 (BY SIMILARITY).	
FT	DISULFID	1972	2123	BY SIMILARITY.	
FT	DISULFID	1950	2085	BY SIMILARITY.	
FT	DISULFID	1927	2088	BY SIMILARITY.	
FT	DISULFID	1993	2001	BY SIMILARITY.	
FT	DISULFID	2724	2774	BY SIMILARITY.	
FT	DISULFID	2739	2788	BY SIMILARITY.	
FT	DISULFID	2750	2804	BY SIMILARITY.	
FT	DISULFID	2754	2806	BY SIMILARITY.	
FT	DISULFID	?	2811	BY SIMILARITY.	
FT	CARBOHYD	99	99	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	156	156	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	211	211	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	666	666	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	857	857	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1231	1231	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1515	1515	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1574	1574	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2223	2223	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2290	2290	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2357	2357	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2400	2400	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2546	2546	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2585	2585	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2790	2790	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CONFLICT	55	55	G -> E (IN REF. 2).	
FT	CONFLICT	70	70	V -> I (IN REF. 3).	
FT	CONFLICT	266	266	A -> G (IN REF. 2).	
FT	CONFLICT	280	280	I -> V (IN REF. 2).	
FT	CONFLICT	409	411	VCH -> ICQ (IN REF. 2).	
FT	CONFLICT	994	994	G -> A (IN REF. 1).	
FT	CONFLICT	1021	1021	F -> L (IN REF. 2).	
FT	CONFLICT	2381	2381	L -> P (IN REF. 2).	
FT	CONFLICT	2406	2406	P -> L (IN REF. 2).	
SQ	SEQUENCE	2813	AA; 309716 MW; 5DF93E1E5E72F80C	CRC64;	
Query Match					7.8%; Score 76.5; DB 1; Length 2813;
Best Local Similarity					21.3%; Pred. No. 39;
Matches					45; Conservative 29; Mismatches 74; Indels 63; Gaps 8;
Qy	26	TDDRCAPAKEEGMMINIS	---IGYHYPPIC-----LGRAPGCLMPAVQNWLVVEVPT	74	
Db	859	TDHVCDAATCAIGAHLTFDGLKLPFGECQYVLVDYCGSNPGTFRILVNGEGCSYPS	918		
Qy	75	VSPNRSRTYHWVSGMSLRPRVNYLQDSYQBSLKFRTGKTCPEIKPGSKNTEVL	----130		
Db	919	VCKKRVTI-LVEG-----GEIELFDGEVNVKPKMKDETHFEVVEGQYVILLGLKAL	970		
Qy	131	--VNEECVANSVWT-----LONNEFGTI-----IDLGTSRSLTPQ	163		
Db	971	SVVNDHRLSLSVTLKRTYQVQVCGLCNFGDQIONDFTSSSLQIEEDPVDVFGNSWKVNPQ	1030		
Qy	164	LLRT-----NSVSKTSSCR	181		
Db	1031	CADTKKVPDLSSPAVCHNNIMKQTMVDSSCR	1061		
RESULT					8

CYRB_MOUSE	STANDARD;	PRT;	896	AA.
ID	CYRB_MOUSE			
AC	P26955;			
DT	01-OCT-1993 (Rel. 27, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	CYTOKINE RECEPTOR COMMON BETA CHAIN PRECURSOR.			
GN	CSF2RB OR CSFR2B1 OR AIC2B OR IL3RB1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-90319131; PubMed-16953379;			
RA	Gorman D.M., Itoh N., Kitamura T., Schreurs J., Yonehara S.,			
RA	Yahara I., Arai K., Miyajima A.,			
RT	"Cloning and expression of a gene encoding an interleukin 3 receptor-			
RT	like protein: identification of another member of the cytokine			
RT	receptor gene family."			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:5459-5463(1990).			
CC	-1- FUNCTION: HIGH AFFINITY RECEPTOR FOR INTERLEUKIN-3, INTERLEUKIN-5			
CC	AND GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR.			
CC	-1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN. THE BETA			
CC	CHAIN IS COMMON TO THE IL-3, IL-5 AND GM-CSF RECEPTORS.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.			
CC	-1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.			
CC	-1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M34397; AAA37204.1; -			
DR	PIR; A35782; A35782.			
DR	MGI; 1339759; Csf2rb1.			
DR	InterPro; IPR002996; CR1A.			
DR	InterPro; IPR000282; Cytok_rcptor_2.			
DR	InterPro; IPR001777; FN_III.			
DR	InterPro; IPR003531; Hematopo_rcptor_S_F1.			
DR	Pfam; PF00041; fn3; 2.			
DR	PROSITE; PS01355; HEMATOPO_REC_S_F1; 1.			
KW	Receptor; Transmembrane; Glycoprotein; Repeat; Signal.			
FT	SIGNAL	1	22	POTENTIAL.
FT	CHAIN	23	896	CYTOKINE RECEPTOR COMMON BETA CHAIN.
FT	DOMAIN	23	441	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	442	463	POTENTIAL.
FT	DOMAIN	464	896	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	132	241	FIBRONECTIN TYPE-III 1.
FT	DOMAIN	343	440	FIBRONECTIN TYPE-III 2.
FT	DISULFID	39	49	BY SIMILARITY.
FT	DISULFID	77	94	BY SIMILARITY.
FT	CARBOHYD	62	62	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	141	141	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	350	350	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	896	AA; 99111	MW; 8CE16EDFDC07A999 CRC64;
Query Match				
Best Local Similarity				
Matches				
Query				
Database				
Query				
Database				

Db 466 TLE-ENCPKALTIRIHDSLNNNTLSVGLIPTNSTNTIMDOKNLKMSTPGOMKAOEV 519

	RESULT 11	RESULT 12
P2CD_MOUSE		

2PCD_MOUSE	STANDARD;	PRT;	598 AA.
AC	Q9QZ67.		
AD	20-AUG-2001	(Rel. 40, Created)	
AE	20-AUG-2001	(Rel. 40, Last sequence update)	
AF	20-AUG-2001	(Rel. 40, Last annotation update)	
AG	PROTEIN PHOSPHATASE 2C DELTA ISOFORM (EC 3.1.3.16) (PP2C-DELTA) (P53-INDUCED PROTEIN PHOSPHATASE 1) (PROTEIN PHOSPHATASE MAGNESIUM-DEPENDENT 1 DELTA).		
AH	PPMID OR WIP1.		
AI	Mus musculus (Mouse).		
AJ	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AK	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
AL	NCBI_TaxID=10090;		
AM	[1]		
AN	SEQUENCE FROM N.A.		
AO	STRAIN=129/SVJ;		
AP	Choi J., Appella E., Donehower L.;		
AQ	"The structure and expression of the murine wild-type p53-induced		
AR	phosphatase 1 (wip1) gene.";		
AS	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.		
AT	- FUNCTION: MIGHT CONTRIBUTE TO GROWTH INHIBITORY PATHWAYS ACTIVATED		
AV	IN RESPONSE TO DNA DAMAGE IN A P53-DEPENDENT MANNER.		
AW	- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O -> A PROTEIN +		
AX	ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).		
AY	- COFACTOR: BINDS TWO MAGNESIUM OR MANGANESE IONS (BY SIMILARITY).		
AZ	- INDUCTION: BY P53.		
BA	- SIMILARITY: BELONGS TO THE PP2C FAMILY.		

- I- SIMILARITY: BELONGS TO THE PP2C FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/or_send_an_email_to_license@isb-sib.ch).

```

CC      DR      EMBL; AF200464; AAF09251.1; -.
CC      DR      MGD; MGI:1858214; Ppmlid.
CC      DR      InterPro; IPR000222; PP2C.
CC      DR      InterPro; IPR003589; PP2C_catalytic.
CC      DR      InterPro; IPR001932; PP2C_domain.
CC      DR      InterPro; IPR003588; PP2C_sig.
CC      DR      Pfam; PF00481; PP2C; 1.
CC      DR      SMART; SM00332; PP2Cc; 1.
CC      DR      SMART; SM00333; PP2C_SIG; 1.
CC      DR      PROSITE; PS01032; PP2C; 1.
CC      DR      Hydrolase; Magnesium; Manganese; Multigene family.
CC      KW      DOMAIN
CC      DR      15 371
CC      DR      PP2C-LIKE.
CC      DR      SEQUENCE 598 AA; 65722 MW; 4DB70B5D48593435 CRC64;

```

Query Match	7.5%;	Score 73.5;	DB 1;	Length 598;
Best Local Similarity	22.1%;	Pred. No. 12;		
Matches 52;	Conservative 26;	Mismatches 86;	Indels 71;	Gaps 10;
QY	3	TPVTWMDNPTEVYV--NDSW--VPGPTDDRCAPKPEEGRGMINISIGYHYHPICLCRA	57	
Db	289	TSVHTLDPRKHVYILGSDGLWNMP-PDAISMCDQEEKYKLMGCEQGQSCAKMLVNRA	347	
QY	58	PG-----CLMPAVONW-----LVEVPYSPNSRRTYHMVSG	88	
Db	348	LGRWRORMLRADNTSAIVCISPEVDNQGFTNEDEFLNLTOSPYY--NSQETCVMTSS	405	
QY	89	MSLRPRVNYLDQFSYO-----RSLKFRPKGKTCPEIPKG-----SKNTE	128	
Db	406	PSSTPIKSPPEEDAMPRLSKDHIPALVRNASEFLEVPABIRAGNIQTVVMTSKDSE	465	
QY	129	VLWWEBCV-----ANSVYILQNNFEFTIIDLGTSRSLIPQLLRNSV	170	

RESULT	12
SRB9_YEAST	

ID	SRB9_YEAST	STANDARD;	PRT;	1420 AA.
AC	P38931.			
DT	01-FEB-1995	(Rel. 31, Created)		
DT	01-FEB-1996	(Rel. 33, Last sequence update)		
DT	01-OCT-1996	(Rel. 34, Last annotation update)		
DE	SUPPRESSOR OF RNA POLYMERASE B SRB9 (SCAL PROTEIN).			
GN	SRB9 OR SCAL OR SSN2 OR YDR443C.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID:4932;			
RN	[1]	SEQUENCE FROM N.A.		
RC	STRAIN: S288C;			
RP	MEDLINE:95293223; PubMed:7774808;			
RA	Hengartner C.J., Thompson C.M., Zhang J., Chao D.M., Liao S.-M.,			
RA	Koleske A.J., Okamura S., Young R.A.;			
RT	"Association of an activator with an RNA polymerase II holoenzyme.";			
RL	Genes Dev. 9:897-910(1995).			
RP	[2]	SEQUENCE FROM N.A.		
RP	Yuryev A., Corden J.L.;			
RA	Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.			
RL	[3]	SEQUENCE FROM N.A.		
RP	Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E., Berno A.,			
RA	Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,			
RA	Hunkle-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,			
RA	Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,			
RA	Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,			
RA	Winant A., Yellon M., Botstein D., Davis R.W.;			
RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: COMPONENT OF THE RNA POLYMERASE II HOLOENZYME AND THE			
CC	MEDIATOR OF ACTIVATION SUBCOMPLEX.			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/			
CC	or send an email to license@isb-sib.ch).			

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

[illegible]

Query Match	7.5%	Score 73.5;	DB 1;	Length 1420;
Best Local Similarity	23.0%;	Pred. No. 34;		
Matches 47;	Conservative	29;	Mismatches	77;
			Indels	51;
			Gaps	11;

Query Match	7.5%	Score 73.5;	DB 1;	Length 1420;
Best Local Similarity	23.0%;	Pred. No. 34;		
Matches 47;	Conservative	29;	Mismatches	77;
Indels	51;	Gaps	11;	

OS Thermoanaerobacter thermohydrosulfuricus (Clostridium
thermohydrosulfuricum).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Thermoanaerobacter.
OX NCBI_TaxID=1516;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E101-69;
RX MEDLINE=90362027; PubMed=2391488;
RA Melasniemi H., Paloheimo M., Hemioe L.;
RT "Nucleotide sequence of the alpha-amylase-pullulanase gene from
RT Clostridium thermohydrosulfuricum.";
RL J. Gen. Microbiol. 136:447-454(1990).
RN [2]
RP SEQUENCE OF 32-39.
RC STRAIN=E101-69;
RX MEDLINE=88268757; PubMed=3260488;
RA Melasniemi H.;
RT "Purification and some properties of the extracellular alpha-amylase-
RT pullulanase produced by Clostridium thermohydrosulfuricum.";
RL Biochem. J. 250:813-818(1988).
RN [3]
RP FIBRONECTIN TYPE III DOMAINS
RX MEDLINE=93028390; PubMed=1409594;
RA Bork P., Doollittle R.F.;
RT "Proposed acquisition of an animal protein domain by bacteria.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:8990-8994(1992).
CC -1- CATALYTIC ACTIVITY: ENDOLYOLYSIS OF 1,4-ALPHA-GLUCOSIDIC
CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.
CC -1- CATALYTIC ACTIVITY: STARCH-DEBRANCHING ENZYME, HYDROLYZES
CC (1-6)-ALPHA-GLUCOSIDIC LINKAGES IN PULLULAN AND STARCH TO
CC FORM MALTOPILOSE.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M28471; AAA23205.1; .
DR PIR; A44765; A44765.
DR InterPro; IPR000461; Alpha-amylase.
DR InterPro; IPR001777; FN_III.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF00041; fn3; 2.
DR SMART; SM00060; FN3; 2.
DR Multifunctional enzyme.
KW Hydrolyase; Glycosidase; Carbohydrate metabolism; Signal; Repeat;
FT SIGNAL 1 31
FT CHAIN 32 1475 AMYLOPULLULANASE.
FT DOMAIN 927 1016 FIBRONECTIN TYPE-III 1.
FT DOMAIN 1163 1255 FIBRONECTIN TYPE-III 2.
FT ACT_SITE 629 629 BY SIMILARITY.
FT ACT_SITE 658 658 BY SIMILARITY.
FT ACT_SITE 735 735 BY SIMILARITY.
SQ SEQUENCE 1475 AA; 165631 MW; 3476C414110BE376 CRC64;

Query Match 7.4%; Score 72.5; DB 1; Length 1475;
Best Local Similarity 28.6%; Pred. No. 44; Mismatches 35; Gaps 9;
Matches 38; Conservative 14;

QY 3 TPVTWMDNPIEVYVNDSDVWPGPTDRCAPAKPEEGMMINISIGYHYPPICLGRAPGLM 62
Db 1031 THVIGVNNPVEVYA--EWAQGLTD-----KPGQGENNI-AQLGYRYIGDVGDA---VY 1079
QY 63 PAVON-----WL-VEVPTVSP---NSRFTYHVMVSGM-----SLRPVNYLQDFS 102

Db 1080 NAVYNKVEGVEISKDWTWVDAQYVGDGSGNNDKYMAKEVPDMVGTWEXIMRESSNQGHWT 1139
QY 103 YQSLKFRPKGKT 115
Db 1140 YTKG----PDGKT 1148
RESULT 15
ENV_MMTVB STANDARD; PRT; 688 AA.
AC P10259;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ENV POLYPROTEIN [CONTAINS: COAT PROTEIN GP52; COAT PROTEIN GP36].
GN ENV,
OS Mouse mammary tumor virus (strain BR6).
OC Viruses; Retroid viruses; Retroviridae; Betaretrovirus.
OX NCBI_TaxID=11758;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87112944; PubMed=3027377;
RA Moore R., Dixon M., Smith R., Peters G., Dickson C.;
RT "Complete nucleotide sequence of a milk-transmitted mouse mammary
RT tumor virus: two frameshift suppression events are required for
RT translation of gag and pol.";
RL J. Virol. 61:480-490(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M15122; AAA46544.1; .
DR PIR; D26795; VCMVM.
KW Coat protein; Polyprotein.
FT PROPEP 1 98 LEADER PEPTIDE.
FT CHAIN 99 474 COAT PROTEIN GP52.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 688 AA; 77176 MW; 46CAAC8D61FFADC2 CRC64;

Query Match 7.4%; Score 72; DB 1; Length 688;
Best Local Similarity 29.1%; Pred. No. 20;
Matches 25; Conservative 10; Mismatches 31; Indels 20; Gaps 5;

QY 4 PVTW-MDNPIEVYVNDSDVWPGPTDRCAPAKPEEGMMINISIGYHYPPICL-- 54
Db 113 PVGWSGTDPIRLVLTNQTMVYLG-----SPDFHGR-NMSGNVHFEKSGDTPICLSF 163
QY 55 --GRAPCLMPAVQVNLVEVPTVSPN 78
Db 164 SFSTPTGCGFQVDRKQVFLSDTPTVDNN 189

Search completed: April 9, 2002, 17:07:05
Job time: 569 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 16:58:14 ; Search time 39.41 Seconds
(without alignments)
349.850 Million cell updates/sec

Title: US-09-490-700-40
Perfect score: 978
Sequence: 1 MVTPTWMDNPIEVYVNDV.....POLLRTNSVSVKCTSESSCR 181

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	768.5	78.6	584	4 VCHUER	retrovirus-related
2	83.5	8.5	990	2 T02309	probable transcrip
3	83	8.5	348	2 T39358	DNA-directed RNA p
4	82	8.4	226	2 A53273	MHC class II histo
5	81	8.3	973	2 J00971	transposase tnpA -
6	80.5	8.2	1003	1 PYVZAM	spheroidin precurs
7	79	8.1	295	2 T00399	probable AP2 domai
8	79	8.1	492	2 T26502	hypothetical prote
9	79	8.1	514	2 T26501	hypothetical prote
10	76	7.8	704	2 T50303	hypothetical prote
11	76	7.8	896	1 A35782	cytokine receptor
12	75.5	7.7	318	2 S76295	hypothetical prote
13	75	7.7	481	2 S04605	glycinin G3 - soyb
14	75	7.7	484	2 S11003	glycinin G3 precur
15	74.5	7.6	602	2 S39782	cyclooxygenase 1 -
16	74.5	7.6	623	2 H96766	unknown protein F2
17	74	7.6	466	2 A36674	transcription fact
18	73.5	7.5	297	2 T36724	probable membrane
19	73.5	7.5	467	2 T21236	hypothetical prote
20	73.5	7.5	1420	2 B57062	SRB9 protein - yea
21	73	7.5	626	2 T04895	vacuolar sorting r
22	73	7.5	1072	2 S76888	hypothetical prote
23	72.5	7.4	365	2 C86647	hypothetical prote
24	72.5	7.4	461	2 T00396	hypothetical prote
25	72.5	7.4	583	2 S64909	probable membrane
26	72.5	7.4	602	2 S69198	prostaglandin G/H
27	72.5	7.4	1475	2 A44765	alpha-amylase [EC
28	72	7.4	285	2 T29490	hypothetical prote
29	72	7.4	341	2 T45350	hypothetical prote

30	72	7.4	688	1 VCMVM	env polyprotein -
31	72	7.4	848	2 C65083	hypothetical prote
32	71.5	7.3	1190	2 S21977	Pm5 protein - huma
33	71.5	7.3	1262	2 T25168	hypothetical prote
34	71.5	7.3	1481	2 S28669	amylolipulanase p
35	71	7.3	251	2 B83413	molybdate-binding
36	71	7.3	387	2 S55164	scn1 protein - fis
37	70.5	7.2	295	2 F83201	conserved hypotet
38	70.5	7.2	533	2 S33701	homeotic protein D
39	70	7.2	256	2 T51150	probable transmem
40	70	7.2	491	2 B86155	probable polygalac
41	70	7.2	570	2 S51404	hypothetical prote
42	70	7.2	688	2 S26388	env polyprotein -
43	69.5	7.1	414	2 T33067	hypothetical prote
44	69.5	7.1	433	2 T39745	hypothetical prote
45	69.5	7.1	488	2 T51502	hypothetical prote

ALIGNMENTS

RESULT 1

VCHUER
retrovirus-related env polyprotein pseudogene - human
C:Species: Homo sapiens (man)
C:Date: 28-Dec-1987 #sequence_revision 04-Jan-1996 #text_change 14-May-1999
C:Accession: E24483
R:Ono, M.; Yasunaga, T.; Miyata, T.; Ushikubo, H.
J. Virol. 60, 589-598, 1986
A:Title: Nucleotide sequence of human endogenous retrovirus genome related to the mou
A:Reference number: A93023; MUID:87036922
A:Accession: E24483
A:Status: conceptual translation of pseudogene
A:Molecule type: DNA
A:Residues: 1-584 <ONO>
A:Cross-references: GB:M14123; NID:gi82227
C:Genetics:
A:Gene: env
C:Keywords: capsid protein; coat protein; polyprotein; pseudogene

Query Match 78.6%; Score 768.5; DB 4; Length 584;
Best Local Similarity 94.7%; Pred. No. 4e-66;
Matches 142; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

Qy 4 PVTWMDNPIEVYVNDVSVVPGPTDDRCAPKEEGMMINISIGYHYPPICLGRAPGCLMP 63
|||||
Db 1 PVTWMDNPIEVYVNDVSVVPGPTDDRCAPKEEGMMINISIGYHYPPICLGRAPGCLMP 60

Qy 64 AVQNWLVETPTSPNSRFTYHMVSGMSLRPRVNYLQDFSQORSKFRPKGTCRKEIPKG 123
|||||
Db 61 AVQNWLVETPTSPISRTYHMVSGMSLRPRVNYLQDFSQORSKFRPKGTCRKEIPKE 120

Qy 124 SKNTEVLWEECVANSVVILQNNFEGTTID 153
|||||
Db 121 SKNTEVLWEECVANSVAIL-NNFEGTTID 149

RESULT 2

T02309
probable transcription elongation factor [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F13P17.5
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 02-Mar-2001
C:Accession: T02309; G84753
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome II BAC F13P17 genomic sequence.
A:Reference number: Z14657
A:Accession: T02309
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-990 <ROU>

A:Cross-references: EMBL:AC004481; NID:g3337347; PID:g3337352

A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A:Reference number: A84420; MUID:20083487

A:Accession: G84753

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-990 <STO>

A:Cross-references: GB:A6002093; NID:g3337352; PIDN:AAC27397.1; GSPDB:GN00139

C:Genetics:

A:Gene: F13P17.5; At2g34210

A:Map position: 2

A:Introns: 97/1; 177/3; 226/3; 280/3; 308/3; 332/2; 451/3; 503/3; 536/2; 552/3; 603/3; 6

C:Keywords: transcription factor

Query Match 8.5%; Score 83.5; DB 2; Length 990;

Best Local Similarity 21.9%; Pred. No. 6.9;

Matches 44; Conservative 41; Mismatches 73; Indels 43; Gaps 10;

Qy 4 PVTWMDNPIEVVNDVW---VPGPTDDRCAPKPEEGMMINISIGYHYPPICL-GRAPG 59

Db 806 PGSGTSPYEATPGDWSGSSTPGRSSYRDAGTP-----INNGFYVLLCLNANAPS 857

Qy 60 CLMPAVQNLVVEP---IVSPNSRFTYHMVSGMSLRPRVNYL-----QDFSQQRSL 107

Db 858 PMTPSSTSLPTPGQAMTGTDLVMSLDIVNIFQFTDVSFFFLCGHHQDGSVVVAL 917

Qy 108 RFRPKGKT-----CPKEIPKSGNTEVLWVECVANSVVIQNNFEFTIIDLGT 156

Db 918 GHRGSETIRATONKVSILVCPK-----NERVKILGGKCGSTAKVIGEDGQDG-IVKLDE 972

Qy 157 SRSLPQLLRNNSVVKCTSE 177

Db 973 SLDI--KILKL-TILAKLVHE 990

RESULT 3

DNA-directed RNA polymerase (EC 2.7.7.6) chain Rpc40 [validated] - fission yeast (Schiz

N:Alternate names: rpa42 protein

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000

C:Accession: T39358; T43505; T50460

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Oliver, K.; Harris, D.

submitted to the EMBL Data Library, March 1999

A:Reference number: Z21848

A:Accession: T39358

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-348 <WOO>

A:Cross-references: EMBL:AL035675; PIDN:CAB38687.1; GSPDB:GN00067; SPDB:SPBCL289.07c

A:Experimental source: strain 972h(-); cosmid c1289

R:Noqi, Y.; Imazawa, Y.

submitted to the EMBL Data Library, December 1998

A:Description: S.pombe rpa42.

A:Reference number: Z22522

A:Accession: T43505

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-348 <NOQ>

A:Cross-references: EMBL:AB021859; PIDN:BAA77385.1

R:Shpakovskii, G.V.; Shematorova, E.K.

Curr. Genet. 36, 208-214, 1999

A:Title: Rpc19 and Rpc40, two alpha-like subunits shared by nuclear RNA polymerases I and

A:Reference number: Z25072

A:Accession: T50460

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-348 <SHP>

A:Cross-references: EMBL:AF082512; PIDN:AAD44503.1

A:Experimental source: strain 972h(-)

C:Genetics:

A:Gene: rpa42; SPBCL289.07c; rpc40+

A:Map position: 2

C:Function:

A:Description: EC 2.7.7.6 [validated; MUID:20009333]

C:Superfamily: Saccharomyces cerevisiae DNA-directed RNA polymerase 40K chain

C:Keywords: nucleotidyltransferase

Query Match 8.5%; Score 83; DB 2; Length 348;

Best Local Similarity 23.5%; Pred. No. 2.2;

Matches 46; Conservative 28; Mismatches 70; Indels 52; Gaps 11;

Qy 13 EYVYNDVSWVP-GPTDDRCAPKPEEGMMINISIGYHYPPICLGR-APGCLMPAVQNLV 70

Db 168 EYSGDLWKPGQROERFADNPIR---VVN-----PDIVVAKLRFG-----QEIDL 211

Qy 71 EVPTV---SPNSRFTYHMVSGMSLRPRVNYLQDFSYSRLKFR---PKGKTCPKPEIKG 123

Db 212 EAHAILGTGQDHAKFSPVATASYRLPTIHLISPIEGEDAVKFKCFKGVIELEGPDG 271

Qy 124 SKNTEV-----LVWEECVANSVVIQNNFEFTIIDLTSRS-----ILPQLLR 166

Db 272 KQARVADVVRKDTVSREC-----LRHPEFADKVQLGRVDRDHYLFSVESTGIMKPDVLF 324

Qy 167 TNSVV---SKCTSESS 179

Db 325 IKSTAVLSKSKCLAVKS 340

RESULT 4

A53273

MHC class II histocompatibility antigen DR alpha chain - horse (fragment)
C:Species: Equus caballus (domestic horse)

C>Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 21-Jan-2000

C:Accession: A53273

R:Albright, D.; Bailey, E.; Woodward, J.G.

Immunogenetics 34, 136-138, 1991

A:Title: Nucleotide sequence of a cDNA clone of the horse (*Equus caballus*) DR alpha gene.

A:Reference number: A53273; MUID:91331619

A:Accession: A53273

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-226 <ALB>

A:Cross-references: GB:M60100; NID:g164236; PIDN:AAA30956.1; PID:g164237

C:Genetics:

A:Gene: DRA

A:Map position: 20

C:Superfamily: class II histocompatibility antigen; immunoglobulin homology

F;97-162/Domain: immunoglobulin homology <IMX>

Query Match

Best Local Similarity 8.4%; Score 82; DB 2; Length 226;

Matches 28; Conservative 16; Mismatches 29; Indels 28; Gaps 5;

Qy 2 VTPVTWMDN--PIEVYVNDVWVPGPTDDRCAPKPEEGMMINISIGYHYPPI----- 52

Db 113 VIDVTWLKNGKPVMTGVSFVFL--PRDDQLFRK-----PHYLPFLPSTEDV 157

Qy 53 --CLGRAPGCLMPAVQNLVVEPTVSPNSRFTYHMVSGMSL 91

Db 158 YDCKVEHWGLDEPLLKHWFEAPT--PLSETENVVCCGLGL 196

RESULT 5

J00971

transposase tnpA - *Enterococcus faecalis* plasmid pAD2 transposon Tn917

N:Alternate names: erythromycin resistance protein, ORF5; erythromycin resistance pro

C:Species: *Enterococcus faecalis*

C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 26-Aug-1999
 C/Accession: JQ0971; E25028; F25028
 R/An, F.Y.; Clewell, D.B.
 Plasmid 25, 121-124, 1991
 A>Title: Tn917 transposase, sequence correction reveals a single open reading frame corr
 A/Reference number: JQ0971; MUID:91312951
 A/Accession: JQ0971
 A/Molecule type: DNA
 A/Residues: 1-973 <ANF>
 A/Cross-references: GB:M11180; GB:M36722; NID:g154957; PIDN:AAA27455.1; PID:g154962
 A/Experimental source: strain DS16
 R/Shaw, J.H.; Clewell, D.B.
 J. Bacteriol. 164, 782-796, 1985
 A/Title: Complete nucleotide sequence of macrolide-lincosamide-streptogramin B-resistan
 A/Reference number: A91808; MUID:86033641
 A/Contents: annotation; erratum
 A/Note: this sequence has been revised in reference JQ0971
 C/Genetics:
 A/Gene: tnpA
 A/Genome: plasmid
 C/Superfamily: transposase Tn21

Query Match 8.3%; Score 81; DB 2; Length 973;
 Best Local Similarity 21.8%; Pred. No. 12;
 Matches 32; Conservative 26; Mismatches 41; Indels 48; Gaps 7;

QY 52 ICLGRAPGLMPAVONWLVPTSPNSRFTYHMVSGSL-----RPRVNY---- 97
 Db 56 LCLARYPGC-----SLSNPIK-----STRLTYSYRQLHDAIDLNSYDHRNTRANHNE 106
 QY 98 -LQDFSVORSLKFRPKGKTCPEIPKSGKNTPEVLWEECVANSVVILQNNF-----GTI 151
 Db 107 ILEFNVHRP-----GSANTQKQLEVLIE---LALENDSDSIYLMKTKI 147
 QY 152 IDLTSRSILPQLRTNSVSKCTSES 178
 Db 148 DFLTRKRIIFPSIATLEDIIISRCRDKA 174

RESULT 6
 PYVZAM
 spheroidin precursor - Amsacta moorei poxvirus
 C/Species: Amsacta moorei poxvirus
 C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
 C/Accession: JQ1436; PQ0289; E41561
 R/Banville, M.; Dumas, F.; Trifiro, S.; Arif, B.; Richardson, C.
 J. Gen. Virol. 73, 559-566, 1992
 A/Title: The predicted amino acid sequence of the spheroidin protein from Amsacta moorei
 A/Reference number: JQ1436; MUID:92185464
 A/Accession: JQ1436
 A/Molecule type: DNA
 A/Residues: 1-1003 <BAN>
 A/Cross-references: GB:M75889; NID:g209629; PIDN:AAA42378.1; PID:g209630
 A/Accession: PQ0289
 A/Molecule type: protein
 A/Residues: 2-7,'X',9-11,'X',13;86-90,'X',92-98,'X',100-102;532-540;728-750;786-802;810-
 R/Hall, R.L.; Moyer, R.W.
 J. Virol. 65, 6516-6527, 1991
 A/Title: Identification, cloning, and sequencing of a fragment of Amsacta moorei entomop
 A/Reference number: A41561; MUID:92046310
 A/Accession: E41561
 A/Molecule type: DNA
 A/Residues: 1-1003 <HAL>
 A/Cross-references: GB:M77182; NID:g209631; PIDN:AAA42383.1; PID:g209636
 C/Comment: This protein is a major component of the occlusion body which serves to prote
 C/Superfamily: Amsacta moorei poxvirus spheroidin
 C/Keywords: acetylated amino end; glycoprotein; leucine zipper
 F:2-1003/Product: spheroidin #status experimental <MAT>
 F:119-140/Region: leucine zipper motif
 F:2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental
 F:47,109,204,210,259,510,529,547,561,799,901,922,955/Binding site: carbohydrate (Asn) (c
 F:537,799/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 8.2%; Score 80.5; DB 1; Length 1003;
 Best Local Similarity 24.3%; Pred. No. 14;
 Matches 27; Conservative 16; Mismatches 39; Indels 29; Gaps 5;

QY 60 CLMPAVQ-----NWLVEVPTSPNSRFTYHMVSG---MSLRPRVNYLQDFSYQSRSLKFR 110
 Db 441 CLKFKPKVKNRLWCILOCDT---SRFIKHMADSGDDLDLDVRLN----- 482
 QY 111 PKGKTCPE-IPKSGKNTPEVLWEECVANSVVILQNNFEGTIIDLGTSRSI 160
 Db 483 -RNDICLKQAIAKHQHTYVILLEYANTYPNCTLSLGNRNRFNVFDMNDKTKI 532

RESULT 7
 T00399
 probable AP2 domain transcription factor [imported] - Arabidopsis thaliana
 N/Alternate names: transcription factor TINY homolog T13E15.5
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Feb-2001
 C/Accession: T00399; F84884
 R/Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; M
 submitted to the EMBL Data Library, July 1997
 A/Description: Arabidopsis thaliana chromosome II BAC T13E15 genomic sequence.
 A/Reference number: Z14146
 A/Accession: T00399
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-295 <ROU>
 A/Cross-references: EMBL:AC002388; NID:g3420042; PID:g2344890
 R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A/Reference number: A84420; MUID:20083487
 A/Accession: F84884
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-295 <STO>
 A/Cross-references: GB:AE002093; NID:g4895256; PIDN:AA032841.1; GSPDB:GN00139
 C/Genetics:
 A/Gene: T13E15.5; At2g44940
 A/Map position: 2

Query Match 8.1%; Score 79; DB 2; Length 295;
 Best Local Similarity 24.4%; Pred. No. 4.4;
 Matches 29; Conservative 29; Mismatches 45; Indels 16; Gaps 6;

QY 68 WLVEVPTSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSRSLKFRPKGKT-CPKEIPKSGKN 126
 Db 125 WLGYPTAEMAAR-AHDVAALAKGTTAYL---NFPKLAGELPRPVTNSPKDI-QAAAS 178
 QY 127 TEVLWEECVANSVVILQNNFEGTIIDLGTSRSILPQLRTNSVSKCT-----SESSC 180
 Db 179 LAAVNQD-----SYNDVSNSEVAEIEAPRVAVQLFSSDSTTTTTSQOEYSEASC 233

RESULT 8
 T26502
 hypothetical protein Y17G7B.10b - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C/Accession: T26502
 R/Smye, R.
 submitted to the EMBL Data Library, June 1998
 A/Reference number: Z20225
 A/Accession: T26502
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA

A;Residues: 1-492 <WIL>
A;Cross-references: EMBL:AL023828; PIDN:CAA19456.1; GSPDB:GN00020; CESP:Y17G7B.10b
A;Experimental source: clone Y17G7B
C;Genetics:
A;Gene: CESP:Y17G7B.10b
A;Map position: 2
A;Introns: 12/3; 60/1; 124/3; 154/2; 240/2; 298/2; 423/1

Query Match 8.1%; Score 79; DB 2; Length 492;
Best Local Similarity 25.0%; Pred. No. 8.1;
Matches 47; Conservative 23; Mismatches 52; Indels 66; Gaps 12;
Qy 4 PVT--WMDNPIEVYVNDVWVPGPTDDRCAPKEPEGMMINISIGYHYPPPI---CLGRAP 58
Db 272 PATTRWTNP-----RSPAREIEECNLS-----PPVKKRMISEA- 306
Qy 59 GCLMPAVQNLVVEVTPVSPNSRFTYHMYSGMSLRPRVNYLQ-DFSYQSRSLKFRPKGKTCP 117
Db 307 -VEIPEVEKETVELPVALNEVFI-----GESLSRSVYIEIGINDAQMALKOKSGGIT-- 358
Qy 118 KEIPKGSNTEVLVW-----EBCVANSVVLQNNFEFGTIIDLTGTSRSLPQLLR--T 167
Db 359 --ICGTGTS---WNFNINKLTCQVQDLMKIV-----AEHCNLPQIPHGDK 401
Qy 168 NSVVSCKT 175
Db 402 NAVSEICT 409

RESULT 9
T26501
hypothetical protein Y17G7B.10a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
A;Accession: T26501
R;Smye, R.
submitted to the EMBL Data Library, June 1998
A;Reference number: Z20225
A;Accession: T26501
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-514 <WIL>
A;Cross-references: EMBL:AL023828; PIDN:CAA19455.1; GSPDB:GN00020; CESP:Y17G7B.10a
A;Experimental source: clone Y17G7B
C;Genetics:
A;Gene: CESP:Y17G7B.10a
A;Map position: 2
A;Introns: 15/2; 34/3; 82/1; 146/3; 176/2; 262/2; 320/2; 445/1

Query Match 8.1%; Score 79; DB 2; Length 514;
Best Local Similarity 25.0%; Pred. No. 8.6;
Matches 47; Conservative 23; Mismatches 52; Indels 66; Gaps 12;
Qy 4 PVT--WMDNPIEVYVNDVWVPGPTDDRCAPKEPEGMMINISIGYHYPPPI---CLGRAP 58
Db 294 PATTRWTNP-----RSPAREIEECNLS-----PPVKKRMISEA- 328
Qy 59 GCLMPAVQNLVVEVTPVSPNSRFTYHMYSGMSLRPRVNYLQ-DFSYQSRSLKFRPKGKTCP 117
Db 329 -VEIPEVEKETVELPVALNEVFI-----GESLSRSVYIEIGINDAQMALKOKSGGIT-- 380
Qy 118 KEIPKGSNTEVLVW-----EBCVANSVVLQNNFEFGTIIDLTGTSRSLPQLLR--T 167
Db 381 --ICGTGTS---WNFNINKLTCQVQDLMKIV-----AEHCNLPQIPHGDK 423
Qy 168 NSVVSCKT 175
Db 424 NAVSEICT 431

RESULT 10

T50303
hypothetical protein SPAPYUG7.03c [imported] - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C;Accession: T50303
R;McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Saunders, D.; Harris, D.
submitted to the EMBL Data Library, January 2000
A;Reference number: Z25059
A;Accession: T50303
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-704 <MCD>
A;Cross-references: EMBL:AL136521; PIDN:CAB66312.1; GSPDB:GN00066; SPDB:SPAPYUG7.03c
A;Experimental source: strain 972h(-); clone plasmid pYUG7
C;Genetics:
A;Gene: SPDB:SPAPYUG7.03c
A;Map position: 1

Query Match 7.8%; Score 76; DB 2; Length 704;
Best Local Similarity 22.3%; Pred. No. 24;
Matches 47; Conservative 29; Mismatches 57; Indels 78; Gaps 11;
Qy 10 NPIEVYVNDVWVPGPTDDRCAPKEPEGMMINISIGYHYPPICLGRAPGCLMPAVQNWL 69
Db 177 NPARSVCNKLK---NEDTLPAEFEE---VSIS---PPVKL----- 208
Qy 70 VEVPTVSPNS---RFTYHMYSGMS---LRPRVNYLQDFSYQ-----RSUKFRPK----- 112
Db 209 -ELPTHSHNSSDTFTNSIVSVSDMVGLEGISASFGFSEDSSSFQDKTKPPRLSFA 267
Qy 113 ---GKTCPEIKPKGSKNTEVLWEECVANSVVLQNNFEFGTIIDLTGTSRSLPQ----- 163
Db 268 DENRENCRTDIYRSDSIHE---YEEPLTSSITSLDSPH---VLDENAPIPLLPKVVSLPD 321
Qy 164 -----LLRTNSVVSCKTSE 177
Db 322 PRFTNVLSAFDALTRTYLLRQNSKVHVHATSQ 352

RESULT 11
A35782
cytokine receptor common beta chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A35782
R;Gorman, D.M.; Itoh, N.; Kitamura, T.; Schreurs, J.; Yonehara, S.; Yahara, I.; Arai, Proc. Natl. Acad. Sci. U.S.A. 87, 5459-5463, 1990
A;Title: Cloning and expression of a gene encoding an interleukin 3 receptor-like protein
A;Reference number: A35782; MUID:90319131
A;Accession: A35782
A;Molecule type: mRNA
A;Residues: 1-896 <GOR>
A;Cross-references: GB:M34397; NID:g191821; PIDN:AAA37204.1; PID:g309101
C;Comment: Mouse high-affinity IL-5, GM-CSF, and one class of high-affinity IL-3 receptor family: interleukin-3 receptor beta chain; cytokine receptor homology
C;Superfamily: cytokine receptor; duplication; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-896/Product: cytokine receptor common beta chain #status predicted <MAT>
F;23-441/Domain: extracellular #status predicted <EXT>
F;39-235/Domain: cytokine receptor homology <CRS1>
F;253-434/Domain: cytokine receptor homology <CRS2>
F;442-463/Domain: transmembrane #status predicted <TM>
F;464-896/Domain: intracellular #status predicted <INT>

Query Match 7.8%; Score 76; DB 1; Length 896;
Best Local Similarity 25.9%; Pred. No. 32;
Matches 29; Conservative 18; Mismatches 39; Indels 26; Gaps 7;
Qy 22 VPG-PTDDRCAPKEPEGMMINISIG-YHYPPICLGRAPGCLMPAVQNWLVEVPTVSPNS 79
Db 794 IPGEPREEVGPASHPHPEGLLVQQVGDYCFLP---GLGPGSLSPH-----SKPPSPSLCS 845


```

Query Match      7.6%; Score 74.5; DB 2; Length 602;
Best Local Similarity 23.3%; Pred. NO. 28;
Matches 37; Conservative 16; Mismatches 61; Indels 45; Gaps 8;

QY 11 PIEVYVNDVWVPGPTDDRCAPKPEEGMMINISIGYHYPPICLG---RAPGCLMPAVON 67
   | | : | | | | : | : | | | | | | | | | | | | | | | | | | |
Db 21 PPPVLLTDA-GVPSVIPCYYPCNQGVCRFGLD-HYQCDCTRTGYSGPNCIPEIWT 78
   | | : | | | | : | : | | | | | | | | | | | | | | | | | | |
QY 68 WLVEVPTVSP-----NSRFTYHNVSGMSLRPRVNYL----- 98
   | | : | | | | : | : | | | | | | | | | | | | | | | | | | |
Db 79 WLSSLRPSPTHELLTHGYWINEFVNATFIREVLGMGWLTVRSNLIPSPPTYNTAHDY 138
   | | : | | | | : | : | | | | | | | | | | | | | | | | | | |
QY 99 ---QDFS-----YQSLKFRPKGKTCPEIKGSKNTEYL 130
   | | : | | | | : | : | | | | | | | | | | | | | | | | | | |
Db 139 ISWESFSNVSYTRILPSVP--KDCP--TPMGTGKKKOL 173
   | | : | | | | : | : | | | | | | | | | | | | | | | | | | |

```

Search completed: April 9, 2002, 16:58:15
Job time: 249 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 16:57:28 ; Search time 34.73 Seconds
(without alignments)
117.279 Million cell updates/sec

Title: US-09-490-700-40

Perfect score: 978

Sequence: 1 MVTPTWMDNPIEVVNDV.....PQLRTNSVSKTSESSCR 181

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80.5	8.2	1003	1 US-07-991-867B-6	Sequence 6, Appli
2	80.5	8.2	1003	1 US-08-107-755A-6	Sequence 6, Appli
3	80.5	8.2	1003	2 US-08-544-332-6	Sequence 6, Appli
4	80.5	8.2	2813	3 US-08-896-449A-2	Sequence 2, Appli
5	80.5	8.2	2813	3 US-09-132-652-2	Sequence 2, Appli
6	73.5	7.5	1420	2 US-08-540-804-14	Sequence 14, Appli
7	73.5	7.5	1420	2 US-08-218-265-14	Sequence 14, Appli
8	73.5	7.5	1420	3 US-08-521-872-14	Sequence 14, Appli
9	73.5	7.5	1420	4 US-08-590-399-14	Sequence 14, Appli
10	69.5	7.1	440	1 US-07-930-686-12	Sequence 12, Appli
11	69.5	7.1	440	2 US-08-460-998-12	Sequence 12, Appli
12	68.5	7.0	192	1 US-08-086-428B-100	Sequence 100, App
13	68.5	7.0	192	2 US-08-468-570-100	Sequence 100, App
14	68.5	7.0	192	2 US-08-290-665A-100	Sequence 100, App
15	68.5	7.0	192	5 PCT-US95-10398-100	Sequence 100, App
16	68.5	7.0	480	4 US-09-078-173A-25	Sequence 25, Appli
17	68.5	7.0	480	4 US-09-537-357-32	Sequence 32, Appli
18	67.5	6.9	473	3 US-08-857-076-99	Sequence 99, Appli
19	67.5	6.9	484	3 US-09-080-044-7	Sequence 7, Appli
20	67.5	6.9	533	1 US-08-220-151-15	Sequence 15, Appli
21	67.5	6.9	533	1 US-08-413-118-15	Sequence 15, Appli
22	67.5	6.9	533	3 US-08-473-446-15	Sequence 15, Appli
23	67.5	6.9	1572	2 US-08-290-731C-5	Sequence 5, Appli
24	67.5	6.9	1596	3 US-09-356-952-3	Sequence 3, Appli
25	67	6.9	599	1 US-08-391-615-4	Sequence 4, Appli
26	67	6.9	799	2 US-08-700-013B-27	Sequence 27, Appli
27	66.5	6.8	807	1 US-07-862-021B-10	Sequence 10, Appli

28	66.5	6.8	807	1	US-08-313-288B-10	Sequence 10, Appli
29	66.5	6.8	807	5	PCT-US93-03164-10	Sequence 10, Appli
30	66.5	6.8	2409	6	5180808-2	Patent No. 5180808
31	66	6.7	176	1	US-08-726-525-2	Sequence 2, Appli
32	66	6.7	176	2	US-08-487-942-2	Sequence 2, Appli
33	66	6.7	176	2	US-08-726-036A-2	Sequence 2, Appli
34	66	6.7	176	4	US-09-083-516-2	Sequence 2, Appli
35	65.5	6.7	410	4	US-09-258-754-451	Sequence 451, App
36	65.5	6.7	828	1	US-08-261-304-2	Sequence 2, Appli
37	65	6.6	338	1	US-08-218-686-2	Sequence 2, Appli
38	65	6.6	338	3	US-08-460-242-2	Sequence 2, Appli
39	65	6.6	732	1	US-08-317-522A-5	Sequence 5, Appli
40	65	6.6	778	1	US-08-439-818A-5	Sequence 5, Appli
41	65	6.6	778	2	US-08-751-965-5	Sequence 5, Appli
42	65	6.6	778	2	US-08-738-975-5	Sequence 5, Appli
43	65	6.6	778	2	US-08-728-626-5	Sequence 5, Appli
44	65	6.6	778	3	US-08-808-599A-5	Sequence 5, Appli
45	64.5	6.6	192	1	US-08-086-428B-97	Sequence 97, Appli

ALIGNMENTS

RESULT 1
US-07-991-867B-6
; Sequence 6, Application US/07991867B
; Patent No. 5476781
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/991,867B
; FILING DATE: 12-DEC-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFL14.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-991-867B-6

```

Query Match      8.2%; Score 80.5; DB 1; Length 1003;
Best Local Similarity 24.3%; Pred. No.1.2;
Matches 27; Conservative 16; Mismatches 39; Indels 29; Gaps 5;

QY   60 CLMPAVQ-----NWLVEVPTSPNSRFTVMVSG---MSLRPRVVYLQDFSQRSLKFR 110
    ||| | | | : : : | | | | | | | | | | | | | | | | | | | | | |
Db   441 CLRKPVPKNRLWGWIIDCDT---SRFKIMADGSDDLDLVDRLN----- 482
                                     : : : | | | | | | | | | | | | | | | | | |

QY   111 PKGTCPKE-IPKSKNTVELWVECIVANSVYILQNNEFGTIIDLGHRSRI 160
    : : : | | | : | : | | | | | | | | | | | | | | | | | | | | | |
Db   483 -RNDICLKQAIKQHVTNVILEYANTYPNCITSLCNRNFNVFDMNDNKTI 532
                                     : : : | | | | | | | | | | | | | | | | | |

RESULT          2
US-08-107-755A-6
; Sequence 6, Application US/08107755A
; Patent No. 5721352
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/107,755A
; FILING DATE: 19-AUG-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,658
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF114.C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-107-755A-6

```

```

Query Match      8.2%; Score 80.5; DB 1; Length 1003;
Best Local Similarity 24.3%; pred. No. 1.2;
Matches 27; Conservative 16; Mismatches 39; Indels 29; Gaps 5;

Qy 60 CLMPAVQ-----NWLVVEVPTVPSNRFYVHWVG---MSLRPVNVLQDFSYQRLKFR 110
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 CLKPKVPKRLRWGLDCDT---SRFIKHMADGSDDLGLDVLRL----- 482

Qy 111 PKGKTCRPE-IPGSKNTEVLVWEECVANSVVILQNNFEFTIIDLGTSRSI 160
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 483 -RNDICLKQAIKQRYTWVILEYANTVPNCTLSLGNRNFNNVFMNDNNKI 532

```

```

RESULT      3
US-08-544-332-6
; Sequence 6, Application US/08544332
; Patent No. 5935777
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gerard H. Bencen
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/544,332
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/991,867
; FILING DATE: 07-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,755
; FILING DATE: 19-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bencen, Gerard H.
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: UFI14.C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-544-332-6

```

Query Match 8.2%; Score 80.5; DB 2; Length 1003;
Best Local Similarity 24.3%; Pred. No. 1.2;
Matches 27; Conservative 16; Mismatches 39; Indels 29; Gaps 5;

QY	60	CLMPAVO-----NWLVEVETVSPNSRFTYHMVSG---MSLRPRVNVQLQDFSTQRSUKFR	110
DB	441	CLRKPVPKNRLRMGLWIDCOT-----SRFIKHMADGSDDLDLVRLN-----	482
QY	111	PKGTCPEKE-IPAGSKNTEVLVWEECVANSVWILQNNEFTGIIIDLTSSRI	160
DB	483	-RNDICLKOAIKOHYTWIILEVANTYPNCTLSLGNRRNNVFMNDNKTI	532

RESULT 4


```

; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI94-03A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1420 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-540-804-14

Query Match          7.5%; Score 73.5; DB 2; Length 1420;
Best Local Similarity 23.0%; Pred. No. 14;
Matches 47; Conservative 29; Mismatches 77; Indels 51; Gaps 11;

Qy 16 VNSWVWPGPTDDRCAPAKEPEG-----MMINISIGYHYP 50
Db 69 INDDPVPPTPAIEHKPVSPDKIGTADYSKPNLPPHYALFKALRRKIYINLALGSHNK 128

Qy 51 PICLRAPGCL-MPAVQNLVVEV-PTVSPNSRFTYHM-VSGMSLRP--RVNYLQDFSYOR 105
Db 129 LIQFGNA--CISLSCVPNLYVQLEPHLFVNGDLTVSLCAKNNGLVPMKEENLEESFLSKH 186

Qy 106 SLKFRPKG-----KTCPEK--IPKGSKNTEVLVWEECVANSVVIQNNNE---FGTIIDL 154
Db 187 ALYLAPSGIRMHLPASQOGYLITPPKHTLELLTTLTSVSHG-INLQNKKNLKWAVVPDL 245

Qy 155 G-----TSRSILPQLLRNTSVV 171
Db 246 GHLNGHTPTIASYLTPLLEAKKLV 269

```

```

RESULT 7
US-08-218-265-14
; Sequence 14, Application US/08218265
; Patent No. 5922585
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; TITLE OF INVENTION: No. 5922585el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Thereof
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/218,265
; FILING DATE: 25-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI94-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1420 amino acids

```

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-218-265-14

Query Match          7.5%; Score 73.5; DB 2; Length 1420;
Best Local Similarity 23.0%; Pred. No. 14;
Matches 47; Conservative 29; Mismatches 77; Indels 51; Gaps 11;

Qy 16 VNSWVWPGPTDDRCAPAKEPEG-----MMINISIGYHYP 50
Db 69 INDDPVPPTPAIEHKPVSPDKIGTADYSKPNLPPHYALFKALRRKIYINLALGSHNK 128

Qy 51 PICLRAPGCL-MPAVQNLVVEV-PTVSPNSRFTYHM-VSGMSLRP--RVNYLQDFSYOR 105
Db 129 LIQFGNA--CISLSCVPNLYVQLEPHLFVNGDLTVSLCAKNNGLVPMKEENLEESFLSKH 186

Qy 106 SLKFRPKG-----KTCPEK--IPKGSKNTEVLVWEECVANSVVIQNNNE---FGTIIDL 154
Db 187 ALYLAPSGIRMHLPASQOGYLITPPKHTLELLTTLTSVSHG-INLQNKKNLKWAVVPDL 245

Qy 155 G-----TSRSILPQLLRNTSVV 171
Db 246 GHLNGHTPTIASYLTPLLEAKKLV 269

RESULT 8
US-08-521-872-14
; Sequence 14, Application US/08521872
; Patent No. 6015682
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Chao, David M.
; TITLE OF INVENTION: No. 6015682el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Thereof
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/521,872
; FILING DATE: 31-AUG-1995
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,265
; FILING DATE: 25-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI94-03A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1420 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-521-872-14

```


Matches 42; Conservative 25; Mismatches 59; Indels 61; Gaps 11;

Qy 10 NPIEVYVNDVWVPGPTDDRCP-----AKPEEGMMINISIGYHYPPICIGRAPGCLMP 63
 Db 177 NRINHIYD-IGDPTCTDDEDCQCTGCTCKDE-----ALCIPPGYTTVMP 220

Qy 64 AVQNWLEVPVPSNSRTHY-----MVSGMSLRPRVNYLQDFSVQSL-----107
 Db 221 PT-----TEKPTTTPK---IYHPGMCMPENNGMTDEARQMFVDKHNYSRLIAKGLAHN 273

Qy 108 --KRPKGTCPKEIPKSKNTEV---LVW-EECVANSVVLONNEFGTTIDLTGTSRSI 160
 Db 274 LGGFAPKA-----ARMKVSYNCEIEANRVEWAKDCTLGYNVAQNQNGYV-----HSL 324

Qy 161 LPQLLRT 167
 Db 325 LPHINKT 331

RESULT 11

US-08-460-998-12

; Sequence 12, Application US/08460998

; Patent No. 5942413

; GENERAL INFORMATION:

; APPLICANT: Sharp, Phillip J

; APPLICANT: Wagland, Barry M

; APPLICANT: Cobon, Gary S

; TITLE OF INVENTION: Nematode Vaccine

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley and Lardner

; STREET: suite 500, 3000 K Street, NW

; CITY: Washington

; STATE: DC

; COUNTRY: United States of America

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/460,998

; FILING DATE: 05-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/930,686

; FILING DATE: 06-OCT-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: AU PK4486

; FILING DATE: 06-FEB-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/AU92/00040

; FILING DATE: 06-FEB-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Bent, Stephen A

; REGISTRATION NUMBER: 29,768

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 672-5300

; TELEFAX: (202) 672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 440 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-460-998-12

Query Match 7.1%; Score 69.5; DB 2; Length 440;

Best Local Similarity 22.5%; Pred. No. 7.4;

Matches 42; Conservative 25; Mismatches 59; Indels 61; Gaps 11;

Qy 10 NPIEVYVNDVWVPGPTDDRCP-----AKPEEGMMINISIGYHYPPICIGRAPGCLMP 63
 Db 177 NRINHIYD-IGDPTCTDDEDCQCTGCTCKDE-----ALCIPPGYTTVMP 220

Qy 64 AVQNWLEVPVPSNSRTHY-----MVSGMSLRPRVNYLQDFSVQSL-----107
 Db 221 PT-----TEKPTTTPK---IYHPGMCMPENNGMTDEARQMFVDKHNYSRLIAKGLAHN 273

Qy 108 --KRPKGTCPKEIPKSKNTEV---LVW-EECVANSVVLONNEFGTTIDLTGTSRSI 160
 Db 274 LGGFAPKA-----ARMKVSYNCEIEANRVEWAKDCTLGYNVAQNQNGYV-----HSL 324

Qy 161 LPQLLRT 167
 Db 325 LPHINKT 331

RESULT 12

US-08-086-428B-100

; Sequence 100, Application US/08086428B

; Patent No. 5514539

; GENERAL INFORMATION:

; APPLICANT: BUKH, J., MILLER, R.H. AND

; APPLICANT: PORCELL, R.H.

; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE

; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE

; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN

; TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES

; NUMBER OF SEQUENCES: 159

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WORDPERFECT 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/086,428B

; FILING DATE: 29-JUN-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: RICHARD W. BORK

; REGISTRATION NUMBER: 36,459

; REFERENCE/DOCKET NUMBER: 2026-4070

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 758-4800

; TELEFAX: (212) 751-6849

; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 100:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 192 amino acids

; TYPE: amino acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; ORGANISM: homosapiens

; INDIVIDUAL ISOLATE: SA7

US-08-086-428B-100

Query Match 7.0%; Score 68.5; DB 1; Length 192;

Best Local Similarity 24.3%; Pred. No. 2.9;

Matches 41; Conservative 19; Mismatches 58; Indels 51; Gaps 9;

Qy 26 TDDRCAPKEEGMMINISIGYHYPPICIGRAPGCLMPAVON-----WLVEVPTVS-----76
 Db 13 TND-CP-----NSSIVYEADNLIL-HAPGCVPCVRQNNVSRVCRWQITPTLSAPNL 60

	Query Match	7.08;	Score 68.5;	DB 2;	Length 192;
	Best Local Similarity	24.3%;	Pred. No. 2.9;		
	Matches 41;	Conservative 19;	Mismatches 58;	Indels 51;	Gaps 9;
Qy	26	TDDRCAPKEEGMMINISIGYHYPPICLGRAPCLMPAVON----	WLVEVPTVS----	76	
Db	13	TND-CP-----	NSSIYEADNLIL-HAPGCVPCVRNNVSRCWQITPLSAPNL	60	
Qy	77	-----PNSRFTYHMVSGMSLRPNVYLQD-----	FSYORSILKFRPKGKT----	CPKEIPK	122

Query Match 7.0%; Score 68.5; DB 2; Length 192;
Best Local Similarity 24.3%; Pred. NO.2.9;
Matches 41; Conservative 19; Mismatches 58; Indels 51; Gaps 9;

[illegible]

Job time: 248 sec

```
RESULT 15
PCT-US95-10398-100
; Sequence 100, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PORCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 100:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORGANISM: hom sapiens
; INDIVIDUAL ISOLATE: SA7
PCT-US95-10398-100

Query Match 7.0%; Score 68.5; DB 5; Length 192;
Best Local Similarity 24.3%; Pred. No. 2.9;
Matches 41; Conservative 19; Mismatches 58; Indels 51; Gaps 9;

Qy 26 TDDRCPAKEEGMMINISICGYHPPICLGRAPCLMPAVON-----WLVEVPIVS----- 76
Db 13 TND-CP-----NSSIYEADNLIL-HAPGCVPCVRQNNVSRVCWQITPLSAPNL 60
Qy 77 -----PNSRFTYHVMVSGMSLRPNVYLQD-----FSYQSRSLKFRPKGKT-----CPKEIPK 122
Db 61 GAVTAPLERRAVDYLAGGAALCSAL-YVGDACGAVFLVQMFYSRPRQHTTVQDCNCISIYS 119
Qy 123 GSKNTEVLVWEECVANSVILONNEFGTIIDLGTSRSLTPOLLRTNSV 171
Db 120 GHITGHRMAWDMNMNWS-----PTTALVMAQLLRIPQV 153
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 16:57:24 ; Search time 34.73 seconds
(without alignments)
363.500 Million cell updates/sec

Title: US-09-490-700-38

Perfect score: 3011

Sequence: 1 MVTPTVMDNDIEVVNDV.....VCRCTPTAPKKTVTSRTGHC 561

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108.5	3.6	2289	US-09-051-019-2	Sequence 2, Appli
2	93	3.1	568	US-07-862-021B-14	Sequence 14, Appl
3	93	3.1	568	PCT-US93-03164-14	Sequence 14, Appl
4	87.5	2.9	750	US-08-202-841A-2	Sequence 2, Appli
5	87.5	2.9	1138	US-08-323-474-8	Sequence 8, Appli
6	87.5	2.9	1138	US-08-469-537A-98	Sequence 98, Appl
7	87.5	2.9	1138	US-08-220-240A-5	Sequence 5, Appli
8	87	2.9	1075	US-08-993-228-19	Sequence 19, Appl
9	85.5	2.8	603	US-09-198-122-2	Sequence 2, Appli
10	85	2.8	397	US-08-948-997-5	Sequence 5, Appli
11	85	2.8	397	US-09-348-817A-5	Sequence 5, Appli
12	85	2.8	1135	US-08-469-537A-97	Sequence 97, Appl
13	84	2.8	660	US-09-111-085-2	Sequence 2, Appli
14	84	2.8	660	US-09-376-781-5	Sequence 5, Appli
15	84	2.8	1259	US-09-187-049-13	Sequence 13, Appl
16	83.5	2.8	719	US-07-943-843-4	Sequence 4, Appli
17	83.5	2.8	719	US-08-347-003-4	Sequence 4, Appli
18	82.5	2.7	801	US-08-725-012-2	Sequence 2, Appli
19	82	2.7	760	US-08-230-491A-2	Sequence 2, Appli
20	82	2.7	760	US-08-619-280A-2	Sequence 2, Appli
21	82	2.7	760	US-08-940-391-2	Sequence 2, Appli
22	82	2.7	1277	US-08-937-236-6	Sequence 6, Appli
23	82	2.7	1292	US-08-569-214-5	Sequence 5, Appli
24	82	2.7	1292	US-08-569-214-6	Sequence 6, Appli
25	82	2.7	1292	US-08-937-236-5	Sequence 5, Appli
26	81.5	2.7	1295	US-08-569-214-2	Sequence 2, Appli
27	81.5	2.7	2555	US-09-058-489-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1

US-09-051-019-2

; Sequence 2, Application US/09051019

; Patent No. 6103229

; GENERAL INFORMATION:

; APPLICANT: KAHMANN, Regine and QUADBECK-SEEGER, Claudia

; TITLE OF INVENTION: Regulatory gene from Ustilago maydis

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Keil & Weinkauf

; STREET: 1101 Connecticut Avenue

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage

; COMPUTER: IBM AT-compatible, Pentium processor

; OPERATING SYSTEM: Windows 98

; SOFTWARE: Wordperfect version 6.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/051,019

; FILING DATE: 31-MAR-1998

; CLASSIFICATION:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2289 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

US-09-051-019-2

Query Match 3.6%; Score 108.5; DB 3; Length 2289;

Best Local Similarity 17.8%; Pred. No. 0.13;

Matches 118; Conservative 77; Mismatches 195; Indels 273; Gaps 32;

Qy 69 LVEVPTVSP-----NSRFTYHVMVSGMSLRPRVN-----YL 98

Db 464 VLEPTQSPSTVASTRRSARKRSEATSTPASSRNSQLQTS-TPMTPLISRRRKGVSPhL 522

Qy 99 QDFSQYQSLRFRPKGKTCPRE-----IPKSKNTEVLWEECVAN-SVVIQLNNEFGTII 152

Db 523 EADSYLRA---QAGNOAQEQMCEICLRGEGDGNMLLCDECNRGYHMYCLO-----PAL 573

Qy 153 DXAPRGQFYHNCSTQSCPSAQVSP-----AVDSLDLTESL-----DXHKH 193

Db 574 TSIPKSWF-----CPPCLVGTGHDGDFDGTSHLSYTFWQRAFAFRDWMWSKQD 624

Qy 194 KKLOSFYLWEEKEGIST-----PRPKIISPVSGPE-----HPELWRLTVASHHIRI 240

Db 194 KKLOSFYLWEEKEGIST-----PRPKIISPVSGPE-----HPELWRLTVASHHIRI 240

Qy 194 KKLOSFYLWEEKEGIST-----PRPKIISPVSGPE-----HPELWRLTVASHHIRI 240

Db 194 KKLOSFYLWEEKEGIST-----PRPKIISPVSGPE-----HPELWRLTVASHHIRI 240

Qy 194 KKLOSFYLWEEKEGIST-----PRPKIISPVSGPE-----HPELWRLTVASHHIRI 240

Db 194 KKLOSFYLWEEKEGIST-----PRPKIISPVSGPE-----HPELWRLTVASHHIRI 240

Db 625 -----HLWRDSEGLATSDYDPTNGLARRVHGTDLVVSEDDVEREFWR-----V 670
Qy 241 WSGNOTLETRYKKPYTIDLSIL-----TVPLQSLCKPPYMLVGVNIVIKPASQITICEN 296
Db 671 HSQKEVEVEYGADVHSTHGSALPTQETHPLSLYSRDKWNL--NNLPILPGS----- 721
Qy 297 CRLFCIDSTFNWQHRILLVRARBGWIPVSTDRPWEASPSIHILTEILKGLVLRKRFI 356
Db 722 --LLOYIKSDIS-----GMTV-----PW----- 737
Qy 357 FTLIAVINGLAVTATAAAGVALHSSVQSVFNWYKKNSTRLM-----NSQ 404
Db 738 -----IYVGMFEST-----FCHNEDHYTYSINQHWGETKYWGIPGEDAEKFNAM 785
Qy 405 SSIDOKLASQINDL--ROTVIMWGDRLDLEHFOQCOWNTSDFCIT--POIYNESEHH-- 459
Db 786 RKAAPDLFETPLDLLFHLTTWMSPEKLKGGVVRVACQDRANEFVTPPKAYHSGFNHGL 845
Qy 460 -----W-----DMVRRHLQGRE-----DNLTLDISK----- 480
Db 846 NLNEAVNFALPDWIFDDLESVRRYORFRKPAVFSHDOLLITVSQOSQTIETAVMLEAAMQ 905
Qy 481 -----LKEQIFEAASKAHNLV-----PCTEAIAGVA 506
Db 906 EMVDREIAKRNALREIIPDLKEEYDVEDVAESHYICSHCTLFSLYGLQTSPTKDGVAICD 965
Qy 507 DGLANLN--PVTWIKTIR--STMIINLILIVVC-----LFCLLLVCRCTPTAPKKT 553
Db 966 HGFEVCNADAPVKWTLKLFSDQLRSILAKVCERAAVPRNWIQRLKKTALGPTTPPLKT 1025
Qy 554 VTS 556
Db 1026 LRS 1028

RESULT 2
US-07-862-021B-14

; Sequence 14, Application US/07862021B
; Patent No. 5279966
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Aviuh
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,021B
; FILING DATE: 19920405
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 568 amino acids
; TYPE: AMINO ACID

; TOPOLOGY: linear
; MOLECULE TYPE: protein
us-07-862-021B-14

Query Match 3.18; Score 93; DB 1; Length 568;
Best Local Similarity 20.88; Pred. No. 0.54; 148; Indels 142; Gaps 22;
Matches 86; Conservative

Qy 14 VYVNDVSVWPGPTDRCPCAKPEEG-----MMINISIGYHYP-----PICLGRAPGCL 61
Db 203 IYWNWSPW-----SACSSATCDKGRMRORMLKAQLDLSVPCPDQDFECMG--PCGS 254
Qy 62 MPAYQNMLV-EVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSRSLKFRPKGT----- 115
Db 255 DDEASTCMSEWITWSPCS-----ASCGMGIEVRYRVKQPEDGSLCKVPTETEKCIV 309
Qy 116 --CPKEIPKSGKNTVLVWEECVANSVVIQNNFETIIDXAPR-GQFYHNCSCQTSQ 171
Db 310 NEECE---PSSCIVTWEAEWEECSATCRMGMKRRH--RMIKMTPADGSMCKADTFVEKC 364
Qy 172 --PSAQVSPAVDSDLTESLDKHKHKLQSFYLW-EWEE-----KGISTPRPKIISP--- 219
Db 365 MWPECHTIPCWLSP-----WSWSDCSVTCGKGTTRQRMKSPSEL 406
Qy 220 -----VSGPEHPMLRLTVASHIRIWSQNTLETIRYKPFYITIDLSILT 265
Db 407 GDCNEELEKQVEKCMLEPCFISCELTWSY-----WS----- 439
Qy 266 VPLQSLCKPPYMLVGVNIVIKP-----ASQITCENCRLFTCIDSTFNWQHRILLVRA 319
Db 440 -ECNKCCKGHHMIRMITMPEQFGGAVCPETVQKKCKLRKCKQSSGNERHRLKDAREK 498
Qy 320 -----EGMWTPVSTDRPWEASPSIHILTEILK-----GVLNR-----SKRF 355
Db 499 RRSEKIKEDSDGEQYPVCKMKPDTA-----WTECTKCGGGIQRERFMTVKRF 546

RESULT 3

PCT-US93-03164-14
; Sequence 14, Application PC/TUS9303164
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Aviuh
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03164
; FILING DATE: 19930402
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525

```

; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 568 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-03164-14

Query Match      3.1%; Score 93; DB 5; Length 568;
Best Local Similarity 20.9%; Pred. No. 0.54;
Matches 86; Conservative 38; Mismatches 148; Indels 142; Gaps 22;

QY 14 VYVNDVWVPCPTDDRCAPAKPEEG-----MMINISIGYHVP-----PICIGRAPCCL 61
Db 203 IYSNWSFW-----SACSSATCDGKKRMRQMLKAQLDLSVPCPDQDFECMG--PGCS 254

QY 62 MPAVONMLV-EVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKT----- 115
Db 255 DDEASTCMGSEWITWSPCS-----ASCGMGIEVRERYVKQFPEDGSLCKVPTETEKIV 309

QY 116 ---CPKEIPKSGKTEVLVWEECVANSVILQNNFEGTIIDXPGR-QGFYHNCSGQTQSC 171
Db 310 NEECE---PSSCIVTEWAEWEECSATCRMGMKRHH--RMIKMTMPADGSMCKADTTEVERK 364

QY 172 --PSAQVSPAVDSDLTESLDKHKHKLQSFYLV--EWEE-----KGISTPRPKIISP--- 219
Db 365 MPPECHTIPCVLSP-----NSENWSDCVTCGKGITRTQRMKLKSPSEL 406

QY 220 -----VSGPEHPELMRLTVASHHTRIMSGNOTLETRYRKPFYTIIDLSILT 265
Db 407 GDCNEELEKQVERKMLPECPISCELTEWSY-----WS----- 439

QY 266 VPLQSLCKPPVMLVGVNIVKP-----ASQITFCNCRFLTCTIDSTFNQHRLLVYRAR 319
Db 440 -ECNKGSGKHMIHTRMITMEPOFGGAVCPETVQRKCRLLKCKQSSGNERRHLLKDAREK 498

QY 320 -----EGMIPVSTDRPWEASPSIHILTILK-----GVLNR-----SKRF 355
Db 499 RRSEKIKSDSGEQYVCKMKPWTA-----WTCTKFCGGGIGQERWTYVKRF 546

RESULT 4
US-08-202-841A-2
; Sequence 2, Application US/08202841A
; Patent No. 6218596
; GENERAL INFORMATION:
; APPLICANT: Hughes, Stephen H.
; APPLICANT: Sutgrave, Pramod
; APPLICANT: Pursel, Vernon
; TITLE OF INVENTION: Enhancement of Musculature in Animals
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,841A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/620,415
; FILING DATE: 03-DEC-1990
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 07/546,449
; FILING DATE: 02-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/373,864
; FILING DATE: 30-JUN-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Alicea, Hector A.
; REGISTRATION NUMBER: 40,891
; REFERENCE/DOCKET NUMBER: 015280-170300US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 750 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 373
; OTHER INFORMATION: /product= "OTHER"
; OTHER INFORMATION: /note= "Xaa = Trp in c-ski;
; OTHER INFORMATION: Xaa = Arg in v-ski"
; US-08-202-841A-2

Query Match      2.9%; Score 87.5; DB 4; Length 750;
Best Local Similarity 20.1%; Pred. No. 3.4;
Matches 46; Conservative 32; Mismatches 82; Indels 69; Gaps 10;

QY 47 YHYPICIGRAPGCLMPAVQN-----WLVEVPTVSPNSRFTYHM----- 85
Db 202 YHE---CFGCKGLLVPELYSNPSAACIOCLDCRLMYPPHKEFVVHSHKSLNRTCHWGFD 258

QY 86 -----VSGMSLRPRVNYLQD-----FSYQSLKFR-PKGKTCPEKPKGSKN 126
Db 259 SANWRSYILLSDQYTGKEKARLQGLLDEMKEKFDYNNKYRKAPRNRSPRVQLRRTKM 318

QY 127 TEVLVWEECVANSVILQNNFEGTIIDXPGRQYHNCSGQTQSCPSAQVSPAVIDS----- 182
Db 319 FKTMLMDP--AGGSAVLQROPDGNVPSDP-----PASKKTKIDDSASQSPASTEKEKO 370

QY 183 -----DLTESLDK-----HKHKKLQSFYLV-----EWEKIGISTPRPKII 217
Db 371 SSXLRSLSSSNKSGICGVHPQRLSAPRPNWSPAVSANEKELSTHLPALI 419

RESULT 5
US-08-323-474-8
; Sequence 8, Application US/08323474
; Patent No. 5447860
; GENERAL INFORMATION:
; APPLICANT: Ziegler, Steven F.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,474
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

```
; APPLICATION NUMBER: US 07/905,600
; FILING DATE: 26-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2609
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; TELEX: 756822
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1138 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-323-474-8

Query Match          2.9%: Score 87.5; DB 1; Length 1138;
Best Local Similarity 19.9%; Pred. No. 7;
Matches 70; Conservative 48; Mismatches 104; Indels 129; Gaps 20;

Qy 72 VPTVSPNSRFTYHVMGSLRPVNYLOD--FSYQSRSLKFRPKGKTCPEKPKGSKNTEV 129
Db 448 VPLAAP--RLLTQKSQLVPLVSFSGDPSTVVR-LHYRPQDST----- 490
Qy 130 LWEECVANSVVLQNNFEFTIDAPR-----GQFYHNCSSG---QTSQSPS 173
Db 491 MDW-----STIVDPSENVTLMNLRPKTGYSVRVOLSRPGEKGAGPPTLMTDCPE 544
Qy 174 AQVSPAVSDLTESLDKHKHKLQSFYLWEEBEGISTPRPKIISPVSQGPPEHPPELWRLTV 233
Db 545 PLLQPLWEGHWEGTDLRL-----VNSLPLVPGPLVGD----- 578
Qy 234 ASHHTIRISNGNQLTETR--YRKPFYTDILNSILT-----VPLQSC-----LKPPYML 278
Db 579 -GFLRLWDGTGRQERRENVSSPOARTALLTGLTPTGTHYQLDVQLYHCTLLGPASPP--- 634
Qy 279 VVGNIVIKPA-----SQTITCENCRLFTCIDSTFNWQH-----RILLVRAR 319
Db 635 --AHVLLPPSGPPAPRHLHAQALSDSEIQL-----TWKHPEALPGIPISKYVVEVQVA 684

RESULT 6
US-08-469-537A-98
; Sequence 98, Application US/08469537A
; Patent No. 5843749
; GENERAL INFORMATION:
; APPLICANT: Maissonpierre, et al.
; TITLE OF INVENTION: EHK AND FOR TYROSINE
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill River Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,537A
; FILING DATE: 06-JUN-1995
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 08/406,247
; FILING DATE: 17-MAR-1995
; APPLICATION NUMBER: USSN 08/144,992
; FILING DATE: 28-OCT-1993
; APPLICATION NUMBER: USSN 07/736,559
; FILING DATE: 26-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kempler, Ph.D., Gail M
; REGISTRATION NUMBER: 32,143
; REFERENCE/DOCKET NUMBER: REG 070C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-345-7400
; TELEFAX: 914-345-7721
; TELEX:
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1138 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-469-537A-98

Query Match          2.9%: Score 87.5; DB 2; Length 1138;
Best Local Similarity 19.9%; Pred. No. 7;
Matches 70; Conservative 48; Mismatches 104; Indels 129; Gaps 20;

Qy 72 VPTVSPNSRFTYHVMGSLRPVNYLOD--FSYQSRSLKFRPKGKTCPEKPKGSKNTEV 129
Db 448 VPLAAP--RLLTQKSQLVPLVSFSGDPSTVVR-LHYRPQDST----- 490
Qy 130 LWEECVANSVVLQNNFEFTIDAPR-----GQFYHNCSSG---QTSQSPS 173
Db 491 MDW-----STIVDPSENVTLMNLRPKTGYSVRVOLSRPGEKGAGPPTLMTDCPE 544
Qy 174 AQVSPAVSDLTESLDKHKHKLQSFYLWEEBEGISTPRPKIISPVSQGPPEHPPELWRLTV 233
Db 545 PLLQPLWEGHWEGTDLRL-----VNSLPLVPGPLVGD----- 578
Qy 234 ASHHTIRISNGNQLTETR--YRKPFYTDILNSILT-----VPLQSC-----LKPPYML 278
Db 579 -GFLRLWDGTGRQERRENVSSPOARTALLTGLTPTGTHYQLDVQLYHCTLLGPASPP--- 634
Qy 279 VVGNIVIKPA-----SQTITCENCRLFTCIDSTFNWQH-----RILLVRAR 319
Db 635 --AHVLLPPSGPPAPRHLHAQALSDSEIQL-----TWKHPEALPGIPISKYVVEVQVA 684

RESULT 7
US-08-220-240A-5
; Sequence 5, Application US/08220240A
; Patent No. 5955291
; GENERAL INFORMATION:
; APPLICANT: Alitalo, Kari
; APPLICANT: Matikainen, Marja-Terttu
; APPLICANT: Partanen, Juha
; APPLICANT: Makela, Tomi
; APPLICANT: Korhonen, Jaana
; TITLE OF INVENTION: ANTIBODIES RECOGNIZING TIE RECEPTOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
```



```

; COUNTRY: Unites States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/220,240A
; FILING DATE: 29-MAR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FI93/00006
; FILING DATE: 08-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/817,800
; FILING DATE: 09-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/167,453
; FILING DATE: 15-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gass, David A.
; REGISTRATION NUMBER: 38,153
; REFERENCE/DOCKET NUMBER: 29151/31958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1138 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-220-240A-5

```

```

Query Match      2.98; Score 87.5; DB 2; Length 1138;
Best Local Similarity 19.98; Pred. No. 7;
Matches 70; Conservative 48; Mismatches 104; Indels 129; Gaps 20;

Qy 72 VPTVSPNSRFTYHVMGSLRPVNYLQD--FSYQSLKFRPKGKTCPEKPKSKNTEV 129
Db 448 VPLAAP--RLTKGSLRLVSPVLSFGDGPSTVR-LHYRPQDST----- 490
Qy 130 LWEECVANSVVIQNNNEFGRIIDXPAP-----CQFYHNCSG-----OTQSCPS 173
Db 491 MDW-----SFIVDPSENVTLMNLRPKTGYSVRVOLSRPGEGGAWGPTLMTDCPE 544
Qy 174 AQVSPAVSDLTESLDKHKHKLQSFYLMWEWEKGIPTPRKIISPVSGPHEPHELRVTV 233
Db 545 PLLQPLWEGHWVEGTDLRL-----VSNLSPLVPGLVGD----- 578
Qy 234 ASHRIWISGNOTLETR--YRKPEYTDLNSILT-----VPLQSC-----LKPPYML 278
Db 579 -GFULRWGDTGRGQERENNVSSPQARTALLGLTPGTHYQLDVLHYCTLLGPASPP--- 634
Qy 279 VVGNVIRKPA-----SQTITCENCRFTCIDSTFNMQH-----RALLVRAR 319
Db 635 --AHVLLPPSGPPAPRHLHAQALSDSEIQL-----TWKHPEALPCPISKYVVEVQVA 684
Qy 320 EG----MWIPVSDRDPWSPASPSIHLTEILKGVNLSRKRFTFTLIAMTGL 366
Db 685 GGAGDPLWIDV--DRPEETS-----TIIRG-LNASTRYLFRMRASIQGL 725

RESULT 8
US-08-993-228-19
; Sequence 19, Application US/08993228
; Patent No. 5976838
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.

```

```

; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,228
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1075 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-993-228-19

Query Match      2.98; Score 87; DB 2; Length 1075;
Best Local Similarity 18.68; Pred. No. 7.2;
Matches 102; Conservative 70; Mismatches 171; Indels 206; Gaps 28;

Qy 50 PPICLGRAPGCLMPAVONMLVEPTVSPNSRFTYHVMGSLRPVNY-LQDES----- 102
Db 431 PP---DRSKLCV--LQSSYFNSPAAQSQSYECLHKVGNLNTVYKFIKEISSSLGR 485
Qy 103 ---YORSKLFPRKGTCPKEIP-----KGSKNTFVLWEECVANSVWL---- 143
Db 486 NDCYKERSRLKKPHKT-SEEVPPCPTPKRGTSQKQAKNTK-----SSAVPNCGLSYTSA 540
Qy 144 ---ONNEFGTIIDXPARGQFYHNCSGQTQSCPSAOVSPA-----VSDLTES----- 187
Db 541 IEGPQTSNAST--SSLEPCNQSRWNAKLQLOSETSS--SPAFTQSSSVGSDNINSPVPL 597
Qy 188 LDKIKHKKLQSFYLMWEWEKGIPTPRKIISPVSGPHEPHELRVTVASHHRIWISGNO-- 245
Db 598 LSKHKSQKQA-----SSPSH-----VTRNGEVVEAWGSDSEY 630
Qy 246 -TLFTRYRKPFYTTDLNSILTVPLOSLKPPYMLVGVNIVTKPASQITTCENCRFTCID 304
Db 631 LALPSHLKQ-----TEVLAKLENLTG-----LLPQKPRCTI--QN-----ID 667
Qy 305 STFNWQHRIILLVRAREGMWIPVSDRDPWEASPSIHLTEILKGVNLSRKRFTFTLIAMV 364
Db 668 D-----WELSEMNSDSEIPTYHVKKKHTR-----L 693
Qy 365 GLIAVTATAAVAGVALHSSVQS-----VNFVNYWQKNSTRLWNSQ----- 404
Db 694 GRVSPSSSDIAS-SLGESIESGPLSDILSDEESSMPLAGMKKYADEKSERASSSEKNE 752
Qy 405 -----SSIDQKIASQI-----NDLRQTVIMMGD----- 427

```

Db 753 HSATKSALIQKLMQIQHODNYEAIWEXIEGFVNKLDEFIQWLNEAMETTENTWTPPKAEM 812
Qy 428 ---RLDLEHHFQLOCDWNTSDFCITPQIYNSEHH-WDMVRRHQLQGRDNLTLDISKLE 483
Db 813 DDLKYLELTHLSFKL--NVDSHCALKEAVEEGHQLLELIASHKAGLKDMLRMITASQWKE 870
Qy 484 QIFASKAH 492
Db 871 LQRQIKRQH 879
RESULT 9
US-09-198-122-2
; Sequence 2, Application US/09198122
; Patent No. 6180380
; GENERAL INFORMATION:
; APPLICANT: Streibhardt, Klaus; Rubsamen-Waigmann, Helga;
; APPLICANT: Holtrich, Uwe
; TITLE OF INVENTION: CLONING OF A MEMBER OF THE SERINE-
; TITLE OF INVENTION: THREONINE-KINASE FAMILY
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPRUNG HORN KRAMER & WOODS
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
; MEDIUM TYPE: storage
; COMPUTER: NEC Powermate SX-20
; OPERATING SYSTEM: DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09198,122
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/601,014
; FILING DATE: 23-FEB-1996
; APPLICATION NUMBER: PCT/EP94/02863
; FILING DATE: 30-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 4329177
; FILING DATE: 30-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: Bayer 9516-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 603 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-09-198-122-2
Query Match 2.8%; Score 85.5; DB 4; Length 603;
Best Local Similarity 19.2%; Pred. No. 3.9;
Matches 116; Conservative 74; Mismatches 221; Indels 193; Gaps 32;
Qy 29 RCPAKPEEGMMINISIGYHYPPICLGRPCGLMPAQVNLVEVP--TVPSNRSRETYHMV 86
Db 12 RAPADPGAGV-----PGVAAPGAPAAAPD-----KEIPEVLVDFRRRY--- 53
Qy 87 SGMSLRPRVNYLQDFSQYRSLKFRP-----KGKTCPEI-----PKGSKNTEVLVWEE 134

Db 54 -----VRGRELGGFAKCFEISDADTKEVFAGKIVKPSLLLLKPHQREKMSMEISIHRS 107
Qy 135 CVANSVW---ILONNEFGTII-----DXAPRGQFYHNCSGQTQSCP 172
Db 108 LAHQHVVGFGHFFEDNDFFVYVLELCRRRSRSLLEPHKRRKALTEPEARYY--LRQIVLGCQ 165
Qy 173 SAQVSPAVDSLDLTS---LDRKHKKKLSQFYL-----WEWEKKGISTPRPKTIIS- VSG 222
Db 166 YLHRNRVTHRLKLGKLNLFNEDLEVKIGDFGLATKVEYDGERKKTLCTGTPNYIAPEVLSK 225
Qy 223 PEHPELWRLTVASHHRIWS-----GNQTLLET-----RYRKPFTYI--DLNS 262
Db 226 KEH-----SPEVDVMSIGCIWYTLVGKPPPETSCLKETYLRKKNEYSIKPHINP 276
Qy 263 ILTVPLQSLCK--PPYMLVGVNIV-----IKPASQITICENC-RLFTCIDSTFNQWH 311
Db 277 VAASLIQKMLQTDPTARTPTINELLNDEFTSGYIPARLPITCLTTPPRFSIAPSLDPSN 336
Qy 312 RILLVRAREGMWIPVSTDRPWEASPSIHLTEILKGLVLRNSKRFFITLIAVIMGLIAVTA 371
Db 337 RKPLTVLNKLENPL--PERPREKEEPV-----VRE 365
Qy 372 TAAVAGVALHSSVQSVNFVYWKNSQSSIDQKLASQINDLRQTVIMWGDRLDL 431
Db 366 TGEVVDCHLSDMLQOLHSVNASKPSERGLVRQEEADPACI-----PIFWVSKWVDY 417
Qy 432 EHHFQL---QCD-----WNTSDFCITPQIYNSEHHWDMVRRHQLQGRDNLTL- 477
Db 418 SDKYGLGVLQCDNSVGVLFNDSTRLI---LYNDGSLQYIER---DGTESYLTVSSHPS 471
Qy 478 ---ISKLK-----EOIFEASKAHLNLPVGPTEAIGVADGLANLNPV-TWIKTIRSTM 525
Db 472 LMKKITLLKFRNYMSEHLL---KAGANITPRE-----GDELARLPYLRTWFRP-RSAI 521
Qy 526 IINL 529
Db 522 ILHL 525
RESULT 10
US-08-948-997-5
; Sequence 5, Application US/08948997
; Patent No. 6008020
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, GREGG
; APPLICANT: COLEMAN, TIM
; TITLE OF INVENTION: BRAIN-ASSOCIATED INHIBITOR OF
; TITLE OF INVENTION: TISSUE-TYPE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,997
; FILING DATE: Oct-10-97
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: A. ANDERS BROOKES
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF336
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 16:58:10 ; Search time 39.41 Seconds
(without alignments)
1084.342 Million cell updates/sec

Title: US-09-490-700-38

Perfect score: 3011

Sequence: 1 MVTPTWMDNPIEVYVNDV.....VCRCTTAPKKTVTSTGHE 561

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2735	90.8	584	4	VCHUER
2	398	13.2	688	1	VCMVMM
3	393	13.1	688	1	VCMVM
4	388	12.9	688	2	S26388
5	340.5	11.3	615	1	VCMVJA
6	152	5.0	584	1	VCMSTA
7	131	4.4	1420	2	T18385
8	131	4.4	1435	2	T18387
9	131	4.4	1463	2	T18386
10	131	4.4	1478	2	T18388
11	125.5	4.2	1407	2	T18381
12	125.5	4.2	1422	2	T18383
13	125.5	4.2	1450	2	T18382
14	125.5	4.2	1465	2	T18384
15	118	3.9	1420	2	T17158
16	118	3.9	1435	2	T46611
17	118	3.9	1452	2	T17157
18	118	3.9	1463	2	T17159
19	118	3.9	1467	2	T17160
20	118	3.9	1478	2	T17185
21	118	3.9	1487	2	T14324
22	116.5	3.9	984	1	VCLJCE
23	116	3.9	1354	2	T18375
24	116	3.9	1369	2	T18379
25	116	3.9	1397	2	T18377
26	116	3.9	1412	2	T18380
27	111.5	3.7	1466	2	T17138
28	111.5	3.7	1471	2	T17149
29	111.5	3.7	1510	2	T17145

30	111.5	3.7	1515	2	T17156
31	110.5	3.7	1341	2	T18301
32	110.5	3.7	1356	2	T18367
33	110.5	3.7	1384	2	T18366
34	110.5	3.7	1399	2	T18370
35	108	3.6	859	1	VCLJE3
36	107.5	3.6	1467	2	T18411
37	107.5	3.6	1472	2	T18413
38	107	3.6	859	1	VCLJEV
39	107	3.6	859	1	VCLJ22
40	107	3.6	859	1	VCLJEM
41	105	3.5	859	1	VCLJWS
42	105	3.5	881	1	VCLJG3
43	104	3.5	787	2	S56268
44	104	3.5	2176	2	T13806
45	103.5	3.4	603	1	VCFVER

ALIGNMENTS

RESULT 1

VCHUER

retrovirus-related env polyprotein pseudogene - human

C:Species: Homo sapiens (man)

C:Date: 28-Dec-1987 #sequence_revision 04-Jan-1996 #text_change 14-May-1999

C:Accession: E24483

R:Ono, M.; Yasunaga, T.; Miyata, T.; Ushikubo, H.

J. Virol. 60, 589-598, 1986

A:Title: Nucleotide sequence of human endogenous retrovirus genome related to the mouse

A:Reference number: A93023; MUID:87036922

A:Accession: E24483

A:Status: conceptual translation of pseudogene

A:Molecule type: DNA

A:Residues: 1-584 <ON0>

A:Cross-references: GB:M14123; NID:gl82227

C:Genetics:

A:Gene: env

C:Keywords: capsid protein; coat protein; polyprotein; pseudogene

Query Match 90.8%; Score 2735; DB 4; Length 584;
Best Local Similarity 94.1%; Pred. No. 3.8e-214;
Matches 512; Conservative 10; Mismatches 20; Indels 2; Gaps 2;

Qy 4 PVTWMDNPIEVYVNDVSVVWPGTDDRCAPKPEEGMMINISIGYHYPTICLGRAPGCLMP 63

Db 1 PVTWMDNPIEVYVNDVSVVWPGTDDRCAPKPEEGMMINISIGYHYPTICLGRAPGCLMP 60

Qy 64 AVQNLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSQYRSLKFRPKGKTCPEIKPG 123

Db 61 AVQNLVVEPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSQYRSLKFRPKGKTCPEIKPE 120

Qy 124 SKNTEVLVWEECVANSVILNNEFGTTIDXAPRGQFTHNCSGQTQSCPSAQVSPAVSD 183

Db 121 SKNTEVLVWEECVANSVILNNEFGTTIDXAPRGQFTHNCSGQTQSCPSAQVSPAVSD 179

Qy 184 LTESLDKHKHKKLQSFYLWEEKEGISTPRKIIISPVSGPEHPELWRLTVASHHRIWSG 243

Db 180 LTESLDKHKHKKLQSFYLWEEKEGISTPRKIIISPVSGPEHPELWRLTVASHHRIWSG 239

Qy 244 NQTLTRYRKPFYTTIDLSILTVLQSCLPKPPYMLVGVNIVIKPASQITTCNCRLFTCI 303

Db 240 NQTLTRYRKPFYTTIDLSILTVLQSCLPKPPYMLVGVNIVIKPASQITTCNCRLFTCI 299

Qy 304 DSTENWQHRIILLVRAREGMWIPVSDRPEWASPSIHTILTEILKGVNRSKRIFTLIAVI 363

Db 300 DSTENWQHRIILLVRAREGMWIPVSDRPEWASPSIHTILTEILKGVNRSKRIFTLIAVI 359

Qy 364 MGLTAVTATAAGVALHSSVQSVNFVWOKNSTRNLANSOSSIDOKLASQINDLRQTVI 423

Db 360 MGLTAVTATAAGVALHSSVQSVNFVWOKNSTRNLANSOSSIDOKLASQINDLRQTVI 419

CU1BB protein - ra
latrophilin-2, spl
latrophilin-2, spl
latrophilin-2, spl
latrophilin-2 (spl)
env polyprotein pr
latrophilin-1, bra
latrophilin-1, bra
env polyprotein pr
env polyprotein pr
env polyprotein pr
env polyprotein pr
hypothetical prote
toucan gene protei
env polyprotein -

```
Qy 424 WNGDRL-DLEHFFQLQCDWNTSDFCITPOIYNESHHWDMVRRHLQGRDNLTLDISKLK 482
||||| ||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 420 WNGDRLMSLEHFRFQLQCDWNTSDFCITPOIYNESHHWDMVRRHLQGRDNLTLDISKLK 479
||||| ||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 483 EQIFPASKAHLNLPVGTFAIAGVADGLANLNPVTWIKTIRSTMIINLILVCLFCLLLV 542
||||| ||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 480 EQIFPASKAHLNLPVGTFAIAGVADGLANLNPVTWIKTIGTSTIINLILVCLFCLLLV 539
||||| ||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 543 CRCT 546
|||||
Db 540 CRCT 543
|||||

RESULT 2
VCVMVM
env polyprotein - mouse mammary tumor virus
N:Contains: coat protein gp36; coat protein gp52
C:Species: mouse mammary tumor virus, MMTV
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 16-Jul-1999
C:Accession: D26795
J:Moore, R.; Dixon, M.; Smith, R.; Peters, G.; Dickson, C.
J. Virol. 61, 480-490, 1987
A:Title: Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: t
A:Reference number: A93030; MUID:87112944
A:Accession: D26795
A:Molecule type: DNA
A:Residues: 1-688 <MOO>
A:Cross-references: EMBL:M15122; NID:g332127; PIDN:AAA46544.1; PID:g332131
C:Genetics:
A:Gene: env
C:Superfamily: type A retrovirus env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein
F:1-97/Domain: leader peptide #status predicted <LPT>
F:98-456/Product: coat protein gp52 #status predicted <GP1>
F:457-688/Product: coat protein gp36 #status predicted <GP2>
F:127,143,297,498,557/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.2%; Score 398; DB 1; Length 688;
Best Local Similarity 23.7%; Pred. No. 3.4e-24;
Matches 146; Conservative 100; Mismatches 217; Indels 152; Gaps 25;

Qy 4 PVTW-MDNPIEYVNDVWVPGPTDRCAPKEPEGMMINISIGVHP-----PICL-- 54
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 113 PVGWSGTDPIRLVNTQWYLG-----SPDFHGR-NMSGNVHFEKSKDTPICLSF 163
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 55 --GRAPGLMPAVQWNLVEPTVSPNSRFTYHMSGSLRPRVNYLQDSYQSLKFRPK 112
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 SFSPTGCFQVDKQVFLSDTPTVDNN-----KPG 192
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 113 GKTCPEIKPGSKNTEVLWVEECVANSVILQNNFEGTIDXPARGQFYHNC----- 164
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 G-----KDKRMWELWLTTLGNSCA---NTKLVPKKLP--PKYPCQIAFKKDA 239
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 165 --SGTQSCPS-----AQVSPAVD-----SDLT 185
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 FWEDESAPRWLPCAFPDQGVSPKPGALGLLWDFSLSPSPVDQSDQIKSKKNLFGNVT 299
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 186 ESLDKHKHKLQSFY---WEWEKGIPTPRKIIISVSGPHEPHELWRLTVASHIRWS 242
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 PPNVKEVHRWEAGWVEPTWEN---SPKDPNDRDFTALVPHTLFLRVAASRLIL-- 354
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 243 GNQTLERYRPFYIDLSILTVLQSLCKPPYMLVVG-----NIVIKPASQTITCNC 297
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 355 -----KRGCFQHEM-----IPTSACTVTPYAILLGLPOLIDIEKRGSTFHISCS 401
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 298 RLFTCIDSTFNQHRILLVRAREGMWIPVST--DRPW---EASPSHIITLTKGLVNRSK 353
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 402 RLTNCLDSS-AYDAAIIIVKRPVYLLPVDIGDEPWFDDSAIQTRYATDLI-----RAK 455
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 354 RFIFTLIAVGLIATATAAAGVALHSSVQSVNFVWKNSTRLNWSQSSIDOKLAS 413
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 RFVAAIILGIALIAITTSFAVATTALVKEMQTATFVNNLHRNVTALASEQRIIDLKLEA 515
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Qy 414 QINDLRQTVIWMG-DRLDLEHFFQLQCDWNTSDFCITPOIYNESHHWDMVRRHLQGR--R 470
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 516 RLNALEEVVLELQGVADANLKTMRSTRCHANYDFICVTPLPYNATE-NWERTRAHLLGIWN 574
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 471 EDNLTDLTKLKEQIFEASKAHLNLPVGTFAIAGVADGLANLNPVTWIKTIRSTMIINLI 530
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 575 DNEISYNIQELTNLISDMSKQHIDAVDLGLAQSPANGVKALNPLDWTQYF-IFIGVGL 633
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 531 LIVVCLFCLLLVCRC 545
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 634 LLVIVLMIFPIVFC 648
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
VCVMVM
env polyprotein - mouse mammary tumor virus (strain GR)
N:Contains: coat protein gp36; coat protein gp52
C:Species: mouse mammary tumor virus, MMTV
C:Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 04-Dec-1994
C:Accession: A03972
R:Redmond, S.M.S.; Dickson, C.
EMBO J. 2, 125-131, 1983
A:Title: Sequence and expression of the mouse mammary tumour virus env gene.
A:Reference number: A03972
A:Accession: A03972
A:Molecule type: DNA
A:Residues: 1-688 <RED>
C:Genetics:
A:Gene: env
C:Superfamily: type A retrovirus env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein
F:1-98/Domain: leader peptide #status predicted <LPT>
F:99-474/Product: coat protein gp52 #status predicted <GP1>
F:475-688/Product: coat protein gp36 #status predicted <GP2>
F:127,143,297,498,557/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.1%; Score 393; DB 1; Length 688;
Best Local Similarity 23.6%; Pred. No. 8.7e-24;
Matches 145; Conservative 99; Mismatches 219; Indels 152; Gaps 25;

Qy 4 PVTW-MDNPIEYVNDVWVPGPTDRCAPKEPEGMMINISIGVHP-----PIC--- 53
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 113 PVGWSGTDPIRLVNTQWYLG-----SPDFHGR-NMSGNVHFEKSKDTPICFSF 163
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 54 --LGRAPGLMPAVQWNLVEPTVSPNSRFTYHMSGSLRPRVNYLQDSYQSLKFRPK 112
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 SFSPTGCFQVDKQVFLSDTPTVDNN-----KPG 192
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 113 GKTCPEIKPGSKNTEVLWVEECVANSVILQNNFEGTIDXPARGQFYHNC----- 164
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 G-----KDKRMWELWLTTLGNSCA---NTKLVPKKLP--PKYPCQIAFKKDA 239
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 165 --SGTQSCPS-----AQVSPAVD-----SDLT 185
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 FWEDESAPRWLPCAFPDQGVSPKPGALGLLWDFSLSPSPVDQSDQIKSKKNLFGNVT 299
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 186 ESLDKHKHKLQSFY---WEWEKGIPTPRKIIISVSGPHEPHELWRLTVASHIRWS 242
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 PPNVKEVHRWEAGWVEPTWEN---SPKDPNDRDFTALVPHTLFLRVAASRLIL-- 354
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 243 GNQTLERYRPFYIDLSILTVLQSLCKPPYMLVVG-----NIVIKPASQTITCNC 297
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 355 -----KRGCFQHEM-----IPTSACTVTPYAILLGLPOLIDIEKRGSTFHISCS 401
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 298 RLFTCIDSTFNQHRILLVRAREGMWIPVST--DRPW---EASPSHIITLTKGLVNRSK 353
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 402 RLTNCLDSS-AYDAAIIIVKRPVYLLPVDIGDEPWFDDSAIQTRYATDLI-----RAK 455
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 354 RFIFTLIAVGLIATATAAAGVALHSSVQSVNFVWKNSTRLNWSQSSIDOKLAS 413
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 RFVAAIILGIALIAITTSFAVATTALVKEMQTATFVNNLHRNVTALASEQRIIDLKLEA 515
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```
QY 392 YWQKSTRWNSQSIDOKLASQINDLRQTVTWMDRL-DLEHHFQLOCDWNTSDFCTTP 450
DB 416 SLSYNTVKVMGTQEDIDKIEDRLSALYDVVRVLGEQVSINFRMKIQCHANYKVICVTK 475
QY 451 QIYNSEHHDWVRHLOG-REDNLTLDISKLEQIEF---ASKAHLNLPVGTETAIAGV 505
DB 476 KPYNTSDFPDWKVKKHLOGWNTNLSLDLQLHNEIIDENSPKATLN-----IADT 528
QY 506 ADG-LANL-NPVTWIKTIRSTM---INLILVVCGLFLLVCR 544
DB 529 VDNFLQNLFSNPPSLHSLWKTLGIGIFVIIIAIVIFVPCVVR 572
RESULT 6
VCM5IA
N:Alternate names: coat polyprotein
N:Contains: surface protein; transmembrane protein
C:Species: mouse intracisternal A-particle MIAE
A:Note: host Mus musculus (house mouse)
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 01-Mar-1996
C:Accession: A41305
R:Reuss, F.U.; Schaller, H.C.
J. Virol. 65, 5702-5709, 1991
A:Title: cDNA sequence and genomic characterization of intracisternal A-particle-related
A:Reference number: A41305; MUID:92015460
A:Accession: A41305
A:Molecule type: mRNA
A:Residues: 1-584 <R>
A:Cross-references: GB:M73818
A:Note: readthrough of three terminators occurs: UGA between codons for 71-Thr and 72-Ala
C:Comment: This particle is a defective retrovirus.
C:Genetics:
A:Gene: env
C:Superfamily: intracisternal A-particle env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein
F:1-526/Domain: extracellular #status predicted <EXT>
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-584/Product: env polyprotein #status predicted <ENV>
F:26-362/Product: surface protein #status predicted <SUP>
F:26-362/Region: hydrophilic #status predicted
F:359-362/Region: cleavage processing #status predicted
F:363-584/Product: transmembrane protein #status predicted <TMP>
F:364-392/Region: hydrophobic #status predicted
F:527-547/Domain: transmembrane #status predicted <TM1>
F:548-584/Domain: intracellular #status predicted <INT>
F:19,58,77,98,129,140,147,230,276,285,311,319,463,469,481,501/Binding site: carbohydrate
```

```
QY 385 OSVNFVNTWQKSTRWNSQSIDOKLASQINDLRQTVTWMDRLDLEHHF-QLOCDWNT 443
DB 393 SGTK-LNQLSADLADAITVQTSASTKLKGLMILNQCLDLAEQIGVLHQAQLGCRKL 451
QY 444 SDFCITLPQIYNSEHHDWVR---HLOGREDNLTLDISKLEQIEFASKAHLNLPVGT 500
DB 452 EALCITSVOYENFTYAANLSROLSLYLAG-----NWSEREDTEALIAAVLAINSTR 504
QY 501 AIAVGADGLANLPVTWIKT---IRSTMIINLILVVC---LFCLLVC---RCTPTAPK 551
DB 505 MDLSLTEGLS-----SWISSAFSYFEKVGVLFGVATCCGLVWMLVCKLRTQOTRDK 559
QY 552 KTVT 555
DB 560 VVIT 563
RESULT 7
TI8385
latrophilin-2 (splice variant bbae) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: TI8385
R:Matsushita, H.; Lelianaova, V.G.; Ushkaryov, Y.A.
FEBS Lett. 443, 348-352, 1999
A:Title: The latrophilin family: multiply spliced G protein-coupled receptors with di
A:Reference number: Z18869; MUID:99148828
A:Accession: TI8385
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1420 <MAT>
A:Cross-references: EMBL:AF111081; NID:g4164044; PID:g4164045; PID:AAD05317.1
C:Superfamily: alpha-latrototoxin receptor, calcium-independent
C:Keywords: alternative splicing; G protein-coupled receptor
Query Match 4.4%; Score 131; DB 2; Length 1420;
Best Local Similarity 18.8%; Pred. NO. 0.047;
Matches 112; Conservative 83; Mismatches 192; Indels 208; Gaps 32;
QY 70 VEVPTVSPNSRFTYHMVSCMSLRPRNVYL----QDFSQYRSLSKFRPKGKTCPEIKPGSK 125
DB 338 VDVP---FPNQ---YQYIAADVNPDRNOLYVWNNFILLRYSLEFGPPD---PAQVP---- 405
QY 126 NTEVLWEECVANSVVIQLQNEF----GTIIDXAPRGQFYHNCSGQTCSCPSAQVSPAV 180
DB 406 -----TTATVITSSAEMFTYVSTSTTSQKGPMTTAVAGSGSGKTRAPPV 454
QY 181 PSD-----LTESLDKHKHKKLQSFYLWEEKEGISTPRP----- 214
DB 455 STTKIPPVTNIFPLPERFCEALDARGIR-----WPQTRGMVVERPCPKGTRGTASYL 507
QY 215 -KIISFVSGPEPELWRLTVASHHI-----RIWSGNQ-----LETRYRKFFYIDLNS 262
DB 508 CVLSTGTWPKGPDLSNCT--SHWYNQLAQKIRSGENASLANELAKHTKGVPFAGDVSS 565
QY 263 -----ILTVPLQSLKPPYMLVGVNIVIKPASOTITCENCRLFTCIDSTEN---- 308
DB 566 SVRLMEQLVDILDAQLE--LKPSEKDSAGRSYKQKREKTC-RAYLKAIVDTVDNLLRP 623
QY 309 -----WQH-----RILLVRAREGMI-----PVSTRDPEWSPSIHILTEI-- 344
DB 624 EALESWKHNSSEQAHTATMLDLTLEEGAFVLADNLVEPTRVSMPT-----NIVLEVAV 678
QY 345 -----LKG-----VLNRSKRFFITLAVI---MGLTAVTATAAV 375
DB 679 LSTEGQVQDFKPLPGIKGAGSSIQLSANTVQKNSRGLAKLVFIILYRSLGQFLSTENATI 738
QY 376 -----AGVALHSSVQSVNFVNYWKNSTRWNSQSIDOKLASQINDLRQTVTW 425
DB 739 KLGAFIGRNSTIANSHVISVI-----NKESRYV-----LTDVLEF 778
QY 426 GDRDLEHHFQLOCD-WNTSDFCITPQIYNSEHHDWVRHLOGREDNLTLDISKLEQ 484
```

Db 779 LPHIDPDNYFNANCSFWNYSE-----RTMNGYWS-----TQCK---LVDTNKRRTT 822
QY 485 IFEASKAHLNLVPCTEAIGVADGLAN--LNPVTWIKTIRSTMTIINILIVVCLF 537
Db 823 CACSHLTNFAILMAHREIA-YKGVHLLTVITWV-----GIVISLVCLAICIF 871

RESULT 8

T18387
latrophilin-2 (splice variant bbbbe) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18387
R:Matsumita, H.; Lelianaova, V.G.; Ushkaryov, Y.A.
FEBS Lett. 443, 348-352, 1999
A:Title: The latrophilin family: multiply spliced G protein-coupled receptors with different
A:Reference number: Z18869; MUID:99148828
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Accession: T18387
A:Molecule type: mRNA
A:Residues: 1-1435 <MAT>
A:Cross-references: EMBL:AF111083; NID:q4164048; PID:q4164049; PIDN:AAD05319.1
C:Superfamily: alpha-latotoxin receptor, calcium-independent
C:Keywords: alternative splicing; G protein-coupled receptor

Query Match 4.4%; Score 131; DB 2; Length 1435;
Best Local Similarity 18.8%; Pred. No. 0.047;
Matches 112; Conservative 83; Mismatches 192; Indels 208; Gaps 32;

QY 70 VEPTVSPNRFTHYMGSLRPRVNYL-----QDFSYQSLKPRPKGKTCPKPIPKGSK 125
Db 358 VDVP--FPNQ---YQYIAADVNPDRNQLYWNNNFILRYSLEFGPPD---PAQVP--- 405
QY 126 NTEVLWEECVANSVWILQNNF-----GTIIDXAPRGQFYHNGSGQTQSCPSAQVSPAV 180
Db 406 -----TTAVTTSSAEMFKTTVSTTSQKGMSTTVAGSQEGSKGTAKPPAV 454
QY 181 DSD-----LTSLDKHKHKLQSFYLWEEKGISTPRP----- 214
Db 455 STTKIPPVTNIFPLPERFCALDARGIR-----WPQORGMMVERPCPKGTRGTASYL 507
QY 215 -KIISPVSGPEHPELRLWLTVAHHI-----RIWNGNOT-----LETRYKPFYITDLS 262
Db 508 CVLSTGTWPKGPDLSNCT--SHVWVLAQKIRSGENAAASLANELAKHTKGPVAGDVSS 565
QY 263 -----ILTVPLQSLKPPYMLVGNIVKPSQTTTCNCRLFTCIDSTFN----- 308
Db 566 SVRLMEQLVDILDALQOE-LKPSKDSAGRSYNKLOKREKTC-RAYLKAIVDTVDNLLRP 623
QY 309 -----WQH-----RILLVREAGMI-----PVSTDRPWEASPSIHLTEI-- 344
Db 624 EALSWKHMSSEQAHTATMLDTEEGAFVLADNLVEPTRVSMPT-----NIVLEVAV 678
QY 345 -----LKG-----VLNRSKRIFTLIAVI-----MGLIAVTATAAV 375
Db 679 LSTGQVDFKFPGLIGKAGSSIQLSANTVKQNSRNGKLKLVFIYRSLGQFLSTENATI 738
QY 376 -----AGVALHSVQSVNFVWQKNSRLNWSQSSIDQKLASQINDLRQTVIWM 425
Db 739 KLGAFTGRNSTIAVNSHVISVSI-----NKESSRVY-----LTDPLVLT 778
QY 345 -----LKG-----VLNRSKRIFTLIAVI-----MGLIAVTATAAV 375
Db 679 LSTGQVDFKFPGLIGKAGSSIQLSANTVKQNSRNGKLKLVFIYRSLGQFLSTENATI 738
QY 376 -----AGVALHSVQSVNFVWQKNSRLNWSQSSIDQKLASQINDLRQTVIWM 425
Db 739 KLGAFTGRNSTIAVNSHVISVSI-----NKESSRVY-----LTDPLVLT 778
QY 426 GDRLDLHHFQLOCD-WNTSDFCITPQIYNESEHHWDMRRHLLQGRDNLTLDISKLEQ 484
Db 779 LPHIDPDNYFNANCSFWNYSE-----RTMNGYWS-----TQCK---LVDTNKRRTT 822
QY 485 IFEASKAHLNLVPCTEAIGVADGLAN--LNPVTWIKTIRSTMTIINILIVVCLF 537
Db 823 CACSHLTNFAILMAHREIA-YKGVHLLTVITWV-----GIVISLVCLAICIF 871

RESULT 9

T18388
latrophilin-2 (splice variant bbbbf) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18388
R:Matsumita, H.; Lelianaova, V.G.; Ushkaryov, Y.A.
FEBS Lett. 443, 348-352, 1999
A:Title: The latrophilin family: multiply spliced G protein-coupled receptors with different
A:Reference number: Z18869; MUID:99148828

T18386
latrophilin-2 (splice variant bbbaf) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18386
R:Matsumita, H.; Lelianaova, V.G.; Ushkaryov, Y.A.
FEBS Lett. 443, 348-352, 1999
A:Title: The latrophilin family: multiply spliced G protein-coupled receptors with different
A:Reference number: Z18869; MUID:99148828
A:Accession: T18386
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1463 <MAT>
A:Cross-references: EMBL:AF111082; NID:q4164046; PID:q4164047; PIDN:AAD05318.1
C:Superfamily: alpha-latotoxin receptor, calcium-independent
C:Keywords: alternative splicing; G protein-coupled receptor

Query Match 4.4%; Score 131; DB 2; Length 1463;
Best Local Similarity 18.8%; Pred. No. 0.049;
Matches 112; Conservative 83; Mismatches 192; Indels 208; Gaps 32;

QY 70 VEPTVSPNRFTHYMGSLRPRVNYL-----QDFSYQSLKPRPKGKTCPKPIPKGSK 125
Db 358 VDVP--FPNQ---YQYIAADVNPDRNQLYWNNNFILRYSLEFGPPD---PAQVP--- 405
QY 126 NTEVLWEECVANSVWILQNNF-----GTIIDXAPRGQFYHNGSGQTQSCPSAQVSPAV 180
Db 406 -----TTAVTTSSAEMFKTTVSTTSQKGMSTTVAGSQEGSKGTAKPPAV 454
QY 181 DSD-----LTSLDKHKHKLQSFYLWEEKGISTPRP----- 214
Db 455 STTKIPPVTNIFPLPERFCALDARGIR-----WPQORGMMVERPCPKGTRGTASYL 507
QY 215 -KIISPVSGPEHPELRLWLTVAHHI-----RIWNGNOT-----LETRYKPFYITDLS 262
Db 508 CVLSTGTWPKGPDLSNCT--SHVWVLAQKIRSGENAAASLANELAKHTKGPVAGDVSS 565
QY 263 -----ILTVPLQSLKPPYMLVGNIVKPSQTTTCNCRLFTCIDSTFN----- 308
Db 566 SVRLMEQLVDILDALQOE-LKPSKDSAGRSYNKLOKREKTC-RAYLKAIVDTVDNLLRP 623
QY 309 -----WQH-----RILLVREAGMI-----PVSTDRPWEASPSIHLTEI-- 344
Db 624 EALSWKHMSSEQAHTATMLDTEEGAFVLADNLVEPTRVSMPT-----NIVLEVAV 678
QY 345 -----LKG-----VLNRSKRIFTLIAVI-----MGLIAVTATAAV 375
Db 679 LSTGQVDFKFPGLIGKAGSSIQLSANTVKQNSRNGKLKLVFIYRSLGQFLSTENATI 738
QY 376 -----AGVALHSVQSVNFVWQKNSRLNWSQSSIDQKLASQINDLRQTVIWM 425
Db 739 KLGAFTGRNSTIAVNSHVISVSI-----NKESSRVY-----LTDPLVLT 778
QY 426 GDRLDLHHFQLOCD-WNTSDFCITPQIYNESEHHWDMRRHLLQGRDNLTLDISKLEQ 484
Db 779 LPHIDPDNYFNANCSFWNYSE-----RTMNGYWS-----TQCK---LVDTNKRRTT 822
QY 485 IFEASKAHLNLVPCTEAIGVADGLAN--LNPVTWIKTIRSTMTIINILIVVCLF 537
Db 823 CACSHLTNFAILMAHREIA-YKGVHLLTVITWV-----GIVISLVCLAICIF 871

RESULT 10

T18388
latrophilin-2 (splice variant bbbbf) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18388
R:Matsumita, H.; Lelianaova, V.G.; Ushkaryov, Y.A.
FEBS Lett. 443, 348-352, 1999
A:Title: The latrophilin family: multiply spliced G protein-coupled receptors with different
A:Reference number: Z18869; MUID:99148828

A:Accession: T18388
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1478 <MAT>
A:Cross-references: EMBL:AF111084; NID:g4164050; PID:g4164051; PIDN:AAD05320.1
C:Superfamily: alpha-latrototoxin receptor, calcium-independent
C:Keywords: alternative splicing; G protein-coupled receptor

Query Match 4.4%; Score 131; DB 2; Length 1478;
Best Local Similarity 18.8%; Pred. No. 0.049;
Matches 112; Conservative 83; Mismatches 192; Indels 208; Gaps 32;

Qy 70 VEVPTVPSNRSFTYHMVSGMSLRPRVNYL----QDYSYQSLKFRPKGKTCPEKPKGSK 125
Db 358 VDVP--FPNQ---YQYIAADVNPDRNQLYVWNNFILRYSLEFGPPD---PAQVP---- 405

Qy 126 NTEVLWEECVANSVVIQNNNEF-----GTIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 180
Db 406 -----TTAVITSSAEMFKTTVSTTTSQKQPMSTTVAGSQEGSGTKRAPPV 454

Qy 181 DSD-----LTESLDKHKHKQSFYLWEEKGISTPRP----- 214
Db 455 STTKIPPVTNIFPLPERCEALDARGIR-----WPQTQGMVMVERPCPKGTRGTASYL 507

Qy 215 -KIISPVSGPEHPELRLTVASHHI-----RIWGSNQT-----LETYRKPFFYIDLNS 262
Db 508 CVLSTGTWNPKGPDLSNCT--SHWYNQLAQKIRSGENAAASLANELAKHTKGFVAGDVSS 565

Qy 263 -----ILTVPLQSLKPPYMLVGVNIVKIPASQITCENCRLFTCIDSTENWQH- 311
Db 566 SVRLMEQLVDILDAQLOE-LKPSEKDSAGRSYKKAIVDTV--DNLLRPEALES---WKHM 619

Qy 312 -----RILLVRAREGMI-----PVSTDRPWEASPSIHILTEI----- 344
Db 620 NSSEQAHTATMLLDLTLEEGAFVLADNLVEPTRVSMPT-----NIVLEVAVLSTEGQVOD 674

Qy 345 -----LKG-----VLNRSKRFTFTLIAVI---MGLIATVATAAV----- 375
Db 675 FKPLGIGKAGSSIQLSANTVKQNSRGLAKLVFIYRSLGQFLSTENATIKLGADFGR 734

Qy 376 -AGVALHSSVQSVFVNYWQKNSRLWNSQSSIDOKLASQINDLRQTVIWMGDRDLDEHH 434
Db 735 NSTIAVNSHVISVSI---NKESRVI-----LTDPLVLTPLHIDPDNY 774

Qy 435 FOLQCD-WNTSDFCITPQIYNESEHHWDMVRHLOGREDNLPLDISKLEQIFEASKAHL 493
Db 775 FNANCSFNWYSE-----RTMNGYWS-----TQCK--LVDTNKTRTTCACSHLTNF 818

Qy 494 NLVPCTEAIGVADGLAN--LNPVTWIKTIRSTMTIINLILIVVCLF 537
Db 819 AILMAHREIA-YKQGVHELLVTITWV-----GIVISLVCLAICIF 858

RESULT 11
T18381
latrophilin-2 (splice variant bbaae) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18381
R:Matsushita, H.; Lelianaova, V.G.; Ushkaryov, Y.A.
FEBS Lett. 443, 348-352, 1999
A:Title: The latrophilin family: multiply spliced G protein-coupled receptors with different functions
A:Reference number: Z18869; MUID:99148828
A:Accession: T18381
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1407 <MAT>
A:Cross-references: EMBL:AF111077; NID:g4164036; PID:g4164037; PIDN:AAD05313.1
C:Superfamily: alpha-latrototoxin receptor, calcium-independent
C:Keywords: alternative splicing; G protein-coupled receptor

Query Match 4.2%; Score 125.5; DB 2; Length 1407;
Best Local Similarity 18.8%; Pred. No. 0.13;
Matches 110; Conservative 84; Mismatches 189; Indels 203; Gaps 32;

Qy 70 VEVPTVPSNRSFTYHMVSGMSLRPRVNYL----QDYSYQSLKFRPKGKTCPEKPKGSK 125
Db 358 VDVP--FPNQ---YQYIAADVNPDRNQLYVWNNFILRYSLEFGPPD---PAQVP---- 405

Qy 126 NTEVLWEECVANSVVIQNNNEF-----GTIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 180
Db 406 -----TTAVITSSAEMFKTTVSTTTSQKQPMSTTVAGSQEGSGTKRAPPV 454

Qy 181 DSD-----LTESLDKHKHKQSFYLWEEKGISTPRP----- 214
Db 455 STTKIPPVTNIFPLPERCEALDARGIR-----WPQTQGMVMVERPCPKGTRGTASYL 507

Qy 215 -KIISPVSGPEHPELRLTVASHHI-----RIWGSNQT-----LETYRKPFFYIDLNS 262
Db 508 CVLSTGTWNPKGPDLSNCT--SHWYNQLAQKIRSGENAAASLANELAKHTKGFVAGDVSS 565

Qy 263 -----ILTVPLQSLKPPYMLVGVNIVKIPASQITCENCRLFTCIDSTENWQH- 311
Db 566 SVRLMEQLVDILDAQLOE-LKPSEKDSAGRSYKKAIVDTV--DNLLRPEALES---WKHM 619

Qy 312 -----RILLVRAREGMI-----PVSTDRPWEASPSIHILTEI----- 344
Db 620 NSSEQAHTATMLLDLTLEEGAFVLADNLVEPTRVSMPT-----NIVLEVAVLSTEGQVOD 674

Qy 345 -----LKG-----VLNRSKRFTFTLIAVI---MGLIATVATAAV----- 375
Db 675 FKPLGIGKAGSSIQLSANTVKQNSRGLAKLVFIYRSLGQFLSTENATIKLGADFGR 734

Qy 376 -AGVALHSSVQSVFVNYWQKNSRLWNSQSSIDOKLASQINDLRQTVIWMGDRDLDEHH 434
Db 735 NSTIAVNSHVISVSI---NKESRVI-----LTDPLVLTPLHIDPDNY 774

Qy 435 FOLQCD-WNTSDFCITPQIYNESEHHWDMVRHLOGREDNLPLDISKLEQIFEASKAHL 493
Db 775 FNANCSFNWYSE-----RTMNGYWS-----TQCK--LVDTNKTRTTCACSHLTNF 818

Qy 494 NLVPCTEAIGVADGLAN--LNPVTWIKTIRSTMTIINLILIVVCLF 537
Db 819 AILMAHREIA-YKQGVHELLVTITWV-----GIVISLVCLAICIF 858

RESULT 12
T18383
latrophilin-2, splice variant bbaee - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18383
R:Matsushita, H.; Lelianaova, V.G.; Ushkaryov, Y.A.
FEBS Lett. 443, 348-352, 1999
A:Title: The latrophilin family: multiply spliced G protein-coupled receptors with different functions
A:Reference number: Z18869; MUID:99148828
A:Accession: T18383
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1422 <MAT>
A:Cross-references: EMBL:AF111079; NID:g4164040; PID:g4164041; PIDN:AAD05315.1
C:Superfamily: alpha-latrototoxin receptor, calcium-independent
C:Keywords: alternative splicing; G protein-coupled receptor

Query Match 4.2%; Score 125.5; DB 2; Length 1422;
Best Local Similarity 18.8%; Pred. No. 0.13;
Matches 110; Conservative 84; Mismatches 189; Indels 203; Gaps 32;

Qy 70 VEVPTVPSNRSFTYHMVSGMSLRPRVNYL----QDYSYQSLKFRPKGKTCPEKPKGSK 125
Db 358 VDVP--FPNQ---YQYIAADVNPDRNQLYVWNNFILRYSLEFGPPD---PAQVP---- 405

Qy 126 NTEVLWEECVANSVVIQNNNEF-----GTIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 180

```
Db 406 -----TTAVITTSAAEMFKTTVTSTTSQKGMSTTVVAGSQEGSKCTKAPPAV 454
QY 181 DSD -----LTESLDKHKHKLQSFVLEWEEKGISTPRP----- 214
Db 455 STTKIPPVTNIFPLPERFCEALDARGIR-----WPQTQGMVVERPCPKGTRGTASYL 507
QY 215 -KIISPVSGPEHPPELWRLTVASHHI-----RIWSGNOT-----LETRYRKPFYTDLNS 262
Db 508 CVLSTGTWPKGPDLSNCT--SHWVNQAOKIRSGENNAASLANELAKHTKGPVFAGDVSS 565
QY 263 -----ILTVPLQSLCKPPYMLVGVNIVIKPASQITTCENCRLFTCIDSTFNMQH- 311
Db 566 SVRLMEQLVDILDQAQOE-LKPSEKDSAGRSYNKAIVDTV--DNLLRPEALES---WKHM 619
QY 312 -----RILLVRAREGMMI-----PVSTDRPWEASPSIHILTEI----- 344
Db 620 NSSEQAHTATMLLDTEEGAFVLADNLVEPTRVSMPT-----NIVLEVAVLSTEGOVQD 674
QY 345 -----LKG-----VLNRSKRFIETLIAVI---MGLIATVATAAV----- 375
Db 675 FKPLPGIKGAGSSITLSANTVKQNSRNLAKLVFIYRSLGQFLSTENATIKLGADFGR 734
QY 376 -AGVALHSSVQSOFNFVWOKNSTRNLWNSQSSIDQKLASQINDLRQTVIWMGDRDLDEHH 434
Db 735 NSTIAVNSHVISI-----NKESRRV-----LTDPLVFTLPHIDPDNY 774
QY 435 FOLQCD-WNTSDFCITPQIYNESEHHWDMVRRHLOQREDNLTLDISKLEQIFEASKAHL 493
Db 775 FNANCSFWNISE-----RTMNGWS-----TOGCK---LVDTKNTRTTCACSHLTNF 818
QY 494 NLVPGTEAIAGVADGLAN--LNPVTWIKTIRSTMIINILIVVCLF 537
Db 819 AILMAHREIA-YKDGVHELLTVITW-----GIVISLVCLAICIF 858

RESULT 13
Tl8382
latrophilin-2 (splice variant bbaaf) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: Tl8382
R:Matsushita, H.; Lelianaova, V.G.; Ushkaryov, Y.A.
FEBS Lett. 443, 348-352, 1999
A:Title: The latrophilin family: multiply spliced G protein-coupled receptors with diffe
A:Reference number: Z18869; MUID:99148828
A:Accession: Tl8382
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1450 <MAT>
A:Cross-references: EMBL:AF111078; NID:q4164038; PID:q4164039; PIDN:AAD05314.1
C:Superfamily: alpha-latrotoxin receptor, calcium-independent
C:Keywords: alternative splicing; G protein-coupled receptor

Query Match 4.2%; Score 125.5; DB 2; Length 1450;
Best Local Similarity 18.8%; Pred. No. 0.13;
Matches 110; Conservative 84; Mismatches 189; Indels 203; Gaps 32;

QY 70 VEPTVSPNSRFTYHYMYSGLSRPRVNYL-----QDFSYQSRSLKFRPKGKTCPKKEIPKGSK 125
Db 358 VDVP---FPNQ---YQYIAAVDYNPRDNQLYVWNNNFILRYSLEFGPPD---PAQVP----- 405
QY 126 NTEVLVWEECVANSVILQNNF-----GTIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 180
Db 406 -----TTAVITTSAAEMFKTTVTSTTSQKGMSTTVVAGSQEGSKGKTKAPPAV 454
QY 181 DSD-----LTESLDKHKHKLQSFVLEWEEKGISTPRP----- 214
Db 455 STTKIPPVTNIFPLPERFCEALDARGIR-----WPQTQGMVVERPCPKGTRGTASYL 507
QY 215 -KIISPVSGPEHPPELWRLTVASHHI-----RIWSGNOT-----LETRYRKPFYTDLNS 262
Db 508 CVLSTGTWPKGPDLSNCT--SHWVNQAOKIRSGENNAASLANELAKHTKGPVFAGDVSS 565
QY 263 -----ILTVPLQSLCKPPYMLVGVNIVIKPASQITTCENCRLFTCIDSTFNMQH- 311
Db 566 SVRLMEQLVDILDQAQOE-LKPSEKDSAGRSYNKAIVDTV--DNLLRPEALES---WKHM 619
QY 312 -----RILLVRAREGMMI-----PVSTDRPWEASPSIHILTEI----- 344
Db 620 NSSEQAHTATMLLDTEEGAFVLADNLVEPTRVSMPT-----NIVLEVAVLSTEGOVQD 674
```

```
Db 508 CVLSTGTWPKGPDLSNCT--SHWVNQAOKIRSGENNAASLANELAKHTKGPVFAGDVSS 565
QY 263 -----ILTVPLQSLCKPPYMLVGVNIVIKPASQITTCENCRLFTCIDSTFNMQH- 311
Db 566 SVRLMEQLVDILDQAQOE-LKPSEKDSAGRSYNKAIVDTV--DNLLRPEALES---WKHM 619
QY 312 -----RILLVRAREGMMI-----PVSTDRPWEASPSIHILTEI----- 344
Db 620 NSSEQAHTATMLLDTEEGAFVLADNLVEPTRVSMPT-----NIVLEVAVLSTEGOVQD 674
QY 345 -----LKG-----VLNRSKRFIETLIAVI---MGLIATVATAAV----- 375
Db 675 FKPLPGIKGAGSSITLSANTVKQNSRNLAKLVFIYRSLGQFLSTENATIKLGADFGR 734
QY 376 -AGVALHSSVQSOFNFVWOKNSTRNLWNSQSSIDQKLASQINDLRQTVIWMGDRDLDEHH 434
Db 735 NSTIAVNSHVISI-----NKESRRV-----LTDPLVFTLPHIDPDNY 774
QY 435 FOLQCD-WNTSDFCITPQIYNESEHHWDMVRRHLOQREDNLTLDISKLEQIFEASKAHL 493
Db 775 FNANCSFWNISE-----RTMNGWS-----TOGCK---LVDTKNTRTTCACSHLTNF 818
QY 494 NLVPGTEAIAGVADGLAN--LNPVTWIKTIRSTMIINILIVVCLF 537
Db 819 AILMAHREIA-YKDGVHELLTVITW-----GIVISLVCLAICIF 858

RESULT 14
Tl8384
latrophilin-2 (splice variant bbaaf) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: Tl8384
R:Matsushita, H.; Lelianaova, V.G.; Ushkaryov, Y.A.
FEBS Lett. 443, 348-352, 1999
A:Title: The latrophilin family: multiply spliced G protein-coupled receptors with di
A:Reference number: Z18869; MUID:99148828
A:Accession: Tl8384
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1465 <MAT>
A:Cross-references: EMBL:AF111080; NID:q4164042; PID:q4164043; PIDN:AAD05316.1
C:Superfamily: alpha-latrotoxin receptor, calcium-independent
C:Keywords: alternative splicing; G protein-coupled receptor

Query Match 4.2%; Score 125.5; DB 2; Length 1465;
Best Local Similarity 18.8%; Pred. No. 0.14;
Matches 110; Conservative 84; Mismatches 189; Indels 203; Gaps 32;

QY 70 VEPTVSPNSRFTYHYMYSGLSRPRVNYL-----QDFSYQSRSLKFRPKGKTCPKKEIPKGSK 125
Db 358 VDVP---FPNQ---YQYIAAVDYNPRDNQLYVWNNNFILRYSLEFGPPD---PAQVP----- 405
QY 126 NTEVLVWEECVANSVILQNNF-----GTIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 180
Db 406 -----TTAVITTSAAEMFKTTVTSTTSQKGMSTTVVAGSQEGSKGKTKAPPAV 454
QY 181 DSD-----LTESLDKHKHKLQSFVLEWEEKGISTPRP----- 214
Db 455 STTKIPPVTNIFPLPERFCEALDARGIR-----WPQTQGMVVERPCPKGTRGTASYL 507
QY 215 -KIISPVSGPEHPPELWRLTVASHHI-----RIWSGNOT-----LETRYRKPFYTDLNS 262
Db 508 CVLSTGTWPKGPDLSNCT--SHWVNQAOKIRSGENNAASLANELAKHTKGPVFAGDVSS 565
QY 263 -----ILTVPLQSLCKPPYMLVGVNIVIKPASQITTCENCRLFTCIDSTFNMQH- 311
Db 566 SVRLMEQLVDILDQAQOE-LKPSEKDSAGRSYNKAIVDTV--DNLLRPEALES---WKHM 619
QY 312 -----RILLVRAREGMMI-----PVSTDRPWEASPSIHILTEI----- 344
Db 620 NSSEQAHTATMLLDTEEGAFVLADNLVEPTRVSMPT-----NIVLEVAVLSTEGOVQD 674
```

```
QY 345 -----LKG-----VLNRSKRFTLTIYAVI---MGLIATATAAV----- 375
      |||      |||      |||      |||      |||      |||      |||
Db 675 FKPLGIGAGASSIQLSANTVKQNSRNLAKLVFIYRSLSGLFSTENATIKLGADFIGR 734
      |||      |||      |||      |||      |||      |||      |||
QY 376 --AGVALHSSVQSVNFVWQKNSRNLWNSQSSIDOKLASQINDLRQTVWNGDRDLDEHH 434
      :||:| ||| :||:| ||| :||:| ||| :||:| ||| :||:| ||| :||:|
Db 735 NSTIAVNSHVISVI---NKESRVI-----LTDPLVLTPLHPIDPDNY 774
      |||      |||      |||      |||      |||      |||      |||
QY 435 FOLOCD-WNTSDFCITPQIYNESEHHWDMVRRHLQGRDNLTLDISKLEQIFEAASKAHL 493
      |||      |||      |||      |||      |||      |||      |||
Db 775 FNANCSFWNYSE-----RTMNGYWS-----TOGCK---LVDNKTTRTTCACSHLTNF 818
      |||      |||      |||      |||      |||      |||      |||
QY 494 NLVPGTEAIAGVADGLAN--LNPVTWIKTIRSTMIINLILIVVCLF 537
      :||:| ||| :||:| ||| :||:| ||| :||:| ||| :||:| ||| :||:|
Db 819 AILMAHREIA-YKDGVBHLLTITWV-----GIVSLVCLAICIF 858
      |||      |||      |||      |||      |||      |||      |||

RESULT 15
TI17158
CL2AB protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T17158
R:Sugita, S.; Ichtchenko, K.; Khvotchev, M.; Sudhof, T.C.
submitted to the EMBL Data Library, July 1998
A:Description: CL family.
A:Reference number: 218712
A:Accession: T17158
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1420 <SUG>
A:Cross-references: EMBL:AF081149; NID:g3695124; PID:g3695125; PIDN:AAC62655.1
C:Superfamily: alpha-latrototoxin receptor, calcium-independent

Query Match 3.9%; Score 118; DB 2; Length 1420;
Best Local Similarity 19.4%; Pred. No. 0.53;
Matches 115; Conservative 78; Mismatches 195; Indels 206; Gaps 33;

QY 70 VEVPTVPSNSRFTYHWSGMSLRPNVYL---QDFSQVRSIKRPKGTCPKEIPKGSK 125
      |||      |||      |||      |||      |||      |||      |||
Db 358 VDVP--FPNQ---YQYIAADYNPHDNLQYVWNNNFILRYSLEFGPPD---PAQVP---- 405
      |||      |||      |||      |||      |||      |||      |||
QY 126 NTEVLVWEECVANSYVILQNNFEFTIIDXAPRGQFYHNCSGQTQSCPSAQVSPAVDSD-- 183
      |||      |||      |||      |||      |||      |||      |||
Db 406 TTAVTI-----TSSAELEFKTTVSTTSQSGPVSVSVAGPQEGSRGTPPPAVSTTKI 459
      |||      |||      |||      |||      |||      |||      |||
QY 184 -----LIESLDKHKHKLQSYLWEEWEGKISTPRP-----KIIIS 218
      |||      |||      |||      |||      |||      |||      |||
Db 460 PPVTNIFPLPERFCEALEMKGIK-----WFOQRGMVMVERPCPKGTRGTASYLCMAST 512
      |||      |||      |||      |||      |||      |||      |||
QY 219 PVSGPEHPMLRLTVASHHI-----RIWSGNQT-----LETRYRKPFYITDLNS----- 262
      |||      |||      |||      |||      |||      |||      |||
Db 513 GTWNPKGFDLSNCT--SHWVQLAQKIRSGENASLANELAKHTKGTVPAGDVSSSVRLM 570
      |||      |||      |||      |||      |||      |||      |||
QY 263 -----ILTVPLQSLCKPPYMLVGVNIVIPASQITICENCRLFTCIDSTFN----- 308
      |||      |||      |||      |||      |||      |||      |||
Db 571 EQLVDILDAQLE-LKPSEKDSAGRSYNKLOKREKTC--RAYLKAIVDITVDNLLRAETLDC 628
      |||      |||      |||      |||      |||      |||      |||
QY 309 WOH-----RILLVRAREGMWIPVSTDRPWE-----ASPSHILTEI----- 344
      |||      |||      |||      |||      |||      |||      |||
Db 629 WKHMNSSEQAHTATMLDLEEGAF--VLADNLELPTRVSMPTDNIVLEAVLSTEGQVQ 686
      |||      |||      |||      |||      |||      |||      |||
QY 345 -----LKG-----VLNRSKRFTLTIYAVI---MGLIATATAAV----- 375
      |||      |||      |||      |||      |||      |||      |||
Db 687 DFTFHLGFGAFSSQLSANTVKQNSRNLAKLVFIYRSLSGLFSTENATIKLGADLLG 746
      |||      |||      |||      |||      |||      |||      |||
QY 376 --AGVALHSSVQSVNFVWQKNSRNLWNSQSSIDOKLASQINDLRQTVWNGDRDLDEH 433
      :||:| ||| :||:| ||| :||:| ||| :||:| ||| :||:| ||| :||:|
Db 747 RNSTIAVNSHVLSVI---NKESRVI-----LTDPLVLSMPHIDSND 786
      |||      |||      |||      |||      |||      |||      |||
QY 434 HFLOCD-WNTSDFCITPQIYNESEHHWDMVRRHLQGRDNLTLDISKLEQI----- 485
```

```
Db 787 YFNANCSFWNYSE-----RTMNGYWS-----TOGCK---LVDNKTTRTTCACSHLTN 830
      |||      |||      |||      |||      |||      |||      |||
QY 486 FEASKAHLNLVPGTEAIAGVADGLAN--LNPVTWIKTIRSTMIINLILIVVCLF 537
      |||      |||      |||      |||      |||      |||      |||
Db 831 FAILMAHREIV-----YKDGVBHLLTITWV-----GIVSLVCLAICIF 871
      |||      |||      |||      |||      |||      |||      |||

Search completed: April 9, 2002, 16:58:14
Job time: 248 sec
```



```
Db 476 KPYNTSDPPWVKVKKHLOGIWFNTNLSLDLQLNEILDIESPKATLN-----IADT 528
Qy 506 ADG-LANL-NPVTWIKTIRSTMI---INLILVVCFLCULLVCR 544
Db 529 VDNFQNLNFSNPSLSLWKLIGLIGFVIIAIVFVPCVVR 572

RESULT 5
ENV_IPMAE
ID ENV_IPMAE STANDARD; PRT; 584 AA.
AC P31789;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ENV POLYPROTEIN PRECURSOR (COAT POLYPROTEIN) [CONTAINS: SURFACE
DE PROTEIN; TRANSMEMBRANE PROTEIN].
GN ENV.
OS Mouse intracisternal A-particle (IAP-IMAE).
OC Viruses; Retrovirdae; Retroviridae; Intracisternal A-particles.
OX NCBI_TaxID=11932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92015460; PubMed=1920613;
RA Reuss F.U., Schaller H.C.;
RT "cDNA sequence and genomic characterization of intracisternal
RT A-particle-related retroviral elements containing an envelope
RT gene."
RL J. Virol. 65:5702-5709(1991).
CC -I- MISCELLANEOUS: REATHROUGH OF THREE TERMINATORS OCCURS: UGA
CC BETWEEN CODONS FOR 71-THR AND 72-ALA, UGA BETWEEN CODONS FOR
CC 111-HIS AND 112-ARG, AND UAA BETWEEN CODONS FOR 394-VAL AND
CC 395-SER.
CC -I- MISCELLANEOUS: THIS PARTICLE IS A DEFECTIVE RETROVIRUS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M73818; -, NOT_ANNOTATED_CDS.
DR PIR; A41305; VCM5IA.
KW Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.
FT SIGNAL 1 25
FT CHAIN 26 584 ENV POLYPROTEIN.
FT CHAIN 26 362 SURFACE PROTEIN.
FT CHAIN 363 584 TRANSMEMBRANE PROTEIN.
FT TRANSMPM 527 547 POTENTIAL.
FT CARBOHYD 58 58 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 77 77 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 129 129 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 311 311 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 463 463 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 481 481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 584 AA; 65024 MW; 180CCAF8F14A4CD6 CRC64;

Query Match 5.0%; Score 152; DB 1; Length 584;
Best Local Similarity 20.9%; Pred. No. 6.8e-05;
Matches 101; Conservative 74; Mismatches 187; Indels 122; Gaps 24;

Qy 117 PKEIPKSKNTE--VLVWEECVANSVW--ILQNEFG---TIIDXAPRGQF----- 160
```

```
Db 157 PKIAPHCSLEDEGLILPMSDC-QSIIIRWDQSKTFSPNNIDD-PKEFVMMKGLFIQ 214
Qy 161 -----YHN--CSGQOTQSCPSAQSAPVSDSLDTSLDKHKHKKQSFYLWEMEKEGISTP 212
Db 215 DFRMHPFKHWVLCGVNGSC--TELNPLI-----FIOGGAVGKASFTGIS-- 256
Qy 213 RPKIISPVSQPEHPPELWRLTVASH-----HIRWSGNOTLETRYKRPFTYIDLNSIL 264
Db 257 -----REAGYWGIDHSDASQDSYGYTNTSVETGFKTKLVNQINYP----- 295
Qy 265 TVPQSCCLKPPYMLVGVNIVIKPASQTTTCNCRLFTCIDSTFNQWHRILLVPRREGHWI 324
Db 296 STPV--CYPPPELFTLSNDSFEVCSN----DSCWISQCDVTKNTRAMV----ARIPRWI 345
Qy 325 PVSTDRPWEASPSIHLTEILKGLVNLRSKRFTFTLIAVIMGLIAVTATAAAGVALHSSV 384
Db 346 PVPVETPSTLS-----MFKRDRFGTAAIMIIAISASAAAATAGYAMVSAV 392
Qy 385 QSVNFVNTWQKNSTRLNWSQSSIDOKLASQINDLRQTIVIMGDRDLDEHFF-QLQCDWNT 443
Db 393 SGTK-LNQLSADLADAITVQTSASTKLGGLMILNOCLDABEQIGVLHQAQLGCKERKL 451
Qy 444 SDFCITPOIYNESEHHWDMVR---HLOGREDNLTLDISKLEQIFEASKAHLNLVPGTE 500
Db 452 EALCITSQYQYENFTYAANLSHQLSLYLAG-----NWSERFDELTLEALIAAVLKINSTR 504
Qy 501 AIAGVADGLANLNPVTWIKT---IRSTMIINLILIVVC---LFCLLLV--RCPTAPK 551
Db 505 MDLSUTEGLS-----SWISSAFSEYKFWGVGLFGVATCCGLVWMLVCKLRTQOTRDK 559
Qy 552 KTVT 555
Db 560 VVIT 563

RESULT 6
ENV_CAEGV
ID ENV_CAEGV STANDARD; PRT; 942 AA.
AC P31627;
DT 01-JUL-1993 (Rel. 26, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ENV POLYPROTEIN PRECURSOR (COAT POLYPROTEIN) [CONTAINS: SURFACE
DE PROTEIN; TRANSMEMBRANE PROTEIN].
GN ENV.
OS Caprine arthritis encephalitis virus (strain G63) (CAEV).
OC Viruses; Retrovirdae; Retroviridae; Lentiviridae; Lentivirus.
OX NCBI_TaxID=11662;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 81-95.
RX MEDLINE=92015464; PubMed=1656067;
RA Knowles D.P. Jr., Cheevers W.P., McGuire T.C., Brassfield A.L.,
RA Harwood W.G., Stem T.A.;
RT "Structure and genetic variability of envelope glycoproteins of two
RT antigenic variants of caprine arthritis-encephalitis lentivirus."
RL J. Virol. 65:5744-5750(1991).
RN [2]
RP REVISIONS.
RA Knowles D.P.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M60855; AAB88709.2; -
DR PIR; A41307; VCLJ3C6.
DR InterPro; IPR000328; Env_GP41.
```


CC	or send an email to license@isb-sib.ch).	
CC	EMBL; M16575; AAB59863.1; ALT_SEQ.	
DR	PIR; A25610; VCLJEV.	
DR	InterPro; IPR001027; EIAV_GP45.	
DR	InterPro; IPR001361; EIAV_GP90.	
DR	Pfam; PF01045; EIAV_GP45; 1.	
DR	Pfam; PF00971; EIAV_GP90; 1.	
KW	coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.	
FT	SIGNAL 1 22	
FT	CHAIN 23 859	
FT	CHAIN 23 444	
FT	CHAIN 445 859	
FT	TRANSMEM 75 93	
FT	TRANSMEM 446 472	
FT	TRANSMEM 617 636	
FT	TRANSMEM 787 807	
FT	TRANSMEM 816 835	
FT	CARBOHYD 40 40	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 411 411	
FT	CARBOHYD 483 483	
FT	CARBOHYD 490 490	
FT	CARBOHYD 550 550	
FT	CARBOHYD 557 557	
FT	CARBOHYD 752 752	
FT	CARBOHYD 859 859	
SQ	SEQUENCE 859 AA; 97041 MW; BC71EBD00FC5A2B9 CRC64;	
Query Match	3.6%; Score 108; DB 1; Length 859;	
Best Local Similarity	21.0%; Pred. No. 0.51;	
Matches 125; Conservative	83; Mismatches 237; Indels 150; Gaps 35;	
Qy	53 CLGAPGCLMPAVONHVEVPTVSPNSRFTYHVMGSLRPNVYVLPQFSYQSRSLKFRPK 112	
Db	122 CWGSPGC-RPFQNY-----FSVETNRSMHMDNNTATLLE-AYHREITFIYK 166	
Qy	113 GK-----TCPK-BIPKSKNTEVY-----WEECVAN---SVVILQN 145	
Db	167 SCTSDSDHCQEVQCKKVNLSASSNSRVVEDVTNTAEYGFQKWLKLECHQNTENFKTILVPE 226	
Qy	146 NEFGTIIIX---APRGQFYHNC---GOTQSCPQAQVSPAVSDLTESLDKHKHKKLQSF 199	
Db	227 NEVMVINDTWTIPKG-----CNETWARVKRP-----IDLYGIHPIRLCVQPPF 272	
Qy	200 YLWEEKGIS-TPR-----PKIISPVSGPEHPELMRLTVASHHRIWISGNQTLER-YR 252	
Db	273 FL--VOEKGADTSRIGCGPTIFLGLV-EDNKGVVVRGDYACNVNRLNINRKDVTGIYQ 329	
Qy	253 KP-FYITIDLSLITVPLQSLKPPYMLV---GNVIVKPAQQTIC--ENCRFLTCDIST 306	
Db	330 VPFIYCTFTNI-----TSCNSKPIIISVMYETNQV-----QYLLCNNNNNNNYCVQVS 379	
Qy	307 FN--WQHRILLVRREGM-----WIPVSTDRPWEASPSIHILTEI 344	
Db	380 FGVIQAHLELPKRNKRIRNOSFNQYNSINKKTELETWKLVT-----SGVTPLPISSEA 435	
Qy	345 LKGVNLSKRF-IFTLIAVIMGLIAVTAAVAGVALH--SSVQSVNFVNYWQKNSTRLW 401	
Db	436 NTGLIRHKRDFGISAIVAIAAATAAATMSYVALTEVKNIMEVQNHTFEVENST--L 493	
Qy	402 NSQSIDOK---LASQINDLRQTVIWMGDRDLLEHHFOL-QCDWNTSFCITPQYINESE 457	
Db	402 NSQSIDOK---LASQINDLRQTVIWMGDRDLLEHHFOL-QCDWNTSFCITPQYINESE 457	
Db	494 NGMDLIERQIKILYAMILQIOTHADVOLLKERQOVETFNIGCIERTHVFCHT-----G 546	
Qy	458 HMDVMVRRLH--QGREDNLITDISKLKEQIFEASKAHNLVPGTEAIAAGVAGLANLNP- 514	
Db	547 HPWNNSWGLNESTOWDDWVSKMEDLNQELITTLHGARNLAQSMITFWPDSIAQFGKD 606	
Qy	515 -----VTWIKTIRSTMIINLILVVCFLCFLLLVCRCTTPAPK-----KTVTSRTG 559	
Db	607 LWSHIGNNIPGLGAS-IKIYIVMFLIIVLL-----TSSPKILRALMKVTSAG 654	
RESULT 11		
ENV_EIAV9		
ID	ENV_EIAV9 STANDARD; PRT; 859 AA.	
AC	PI1306;	
DT	01-JUL-1989 (Rel. 11, Created)	
DT	01-FEB-1996 (Rel. 33, Last sequence update)	
DT	20-AUG-2001 (Rel. 40, Last annotation update)	
DE	ENV POLYPROTEIN PRECURSOR (COAT POLYPROTEIN) [CONTAINS: COAT PROTEIN GP90; COAT PROTEIN GP45].	
GN	ENV.	
OC	Equine infectious anemia virus (clone 1369) (EIAV).	
OX	Viruses; Retroviral viruses; Retroviridae; Lentiviruses.	
RN	NCBI_TaxID=11670;	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=87236196; PubMed=3035786;	
RA	Kawakami T., Sherman L., Dahlberg J., Gazit A., Yaniv A.,	
RT	Tronick S.R., Aaronson S.A.;	
RT	"Nucleotide sequence analysis of equine infectious anemia virus proviral DNA.";	
RL	Virology 158:300-312(1987).	
RN	[2]	
RP	REVISIONS TO N-TERMINUS.	
RA	Tronick S.R.;	
RL	Submitted (NOV-1987) to the EMBL/GenBank/DBJ databases.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; M16575; AAB59863.1; -	
DR	PIR; C27842; VCLJEV.	
DR	InterPro; IPR001027; EIAV_GP45.	
DR	InterPro; IPR001361; EIAV_GP90.	
DR	Pfam; PF01045; EIAV_GP45; 1.	
DR	Pfam; PF00971; EIAV_GP90; 1.	
KW	Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.	
FT	SIGNAL 1 22	
FT	CHAIN 23 859	
FT	CHAIN 23 444	
FT	CHAIN 445 859	
FT	TRANSMEM 75 93	
FT	TRANSMEM 446 462	
FT	TRANSMEM 614 636	
FT	TRANSMEM 787 807	
FT	TRANSMEM 816 835	
FT	CARBOHYD 40 40	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 186	
FT	CARBOHYD 214 214	
FT	CARBOHYD 233 233	
FT	CARBOHYD 244 244	
FT	CARBOHYD 340 340	
FT	CARBOHYD 368 368	
FT	CARBOHYD 399 399	
FT	CARBOHYD 406 406	
FT	CARBOHYD 112 112	
FT	CARBOHYD 141 141	
FT	CARBOHYD 148 148	
FT	CARBOHYD 186 1	

```

FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 490 490 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 859 AA; 97113 MW; 4B4ED8518CD4F364 CRC64;

Query Match 3.6%; Score 107; DB 1; Length 859;
Best Local Similarity 21.0%; Pred. No. 0.62;
Matches 125; Conservative 82; Mismatches 238; Indels 150; Gaps 35;

QY 53 CLGRAPCCLPAVQNNLVEVTPSPNSRFTYHVMVSGMSLRPRVNYLQDFSYQSLKFRPK 112
DB 122 CWGSPGCG--RPFQNY-----PSYETNRSHMDNNTATLLE-AYHREITFIYK 166
QY 113 GK-----TCPK-EIPKSGKNTVLV-----WEECVAN---SVVILQN 145
DB 167 SSTDSHDCQEQYCKKVNLSNDSNSVRVEDVTNTAEYWGFKWLECNQTFNFTILVPE 226
QY 146 NEFTIIX---APRGQFYHNC---GOTQSCPSAQVSPAVDSLTESLDKHKHKLQSF 199
DB 227 NEWMVINDTWTIPKG-----CNETWARVKRC-----IDILYGIHPILRCVOPPE 272
QY 200 YLWEEKEGIGIS-TPR-----PKIISPVSGPEHPMLRLTVASHIRIWSGNQTLRYK 253
DB 273 FL--VQEKGIADTSRIGNCGPTIFLGLV-EDNKGVVGRDGTACNVRRLNINRKDYGIY 329
QY 253 KP-FYITDLNSILTVPLQSLCKPPYMLV---GNVIKPAQITC---ENCRLETCIDST 306
DB 330 VPIFTCTFTNI-----TSCNNEPIISVIMYETNOV-----QYLLCNNSNNYNCVQSF 379
QY 307 FN--WQHRIILLVRREGM-----WIPVSTDRPWEASPSIHILTEI 344
DB 380 FGIVGAHLELPPNKRIRNQSFNOYCNINNKTELETHKLVK---SGVTPLPISSEA 435
QY 345 LKGVLRNRRF-IFTIIVITMGLIATVATAAGVALH---SSVQSVNFVNYWQKNSTRLW 401
DB 436 NTGLIRKRDGSAIQAATAAATAAATSAATSVYALTEVKNIMEVQNHTEFVNST--L 493
QY 402 NSOSSIDOK---LASOINDLROTVWMDRLDLEHIFQL-OCWNTSDFCITQIYNESE 457
DB 494 NGMDLIERQIKILYAMLOTHADVQLKKEQEQEETFNLCIGCTIERTHVFCHT-----G 546
QY 458 HHWDVRRHL--QGRDNLTLDISKLEQIFEASKAHLNLVPCGTEAIGVADGLANLP- 514
DB 547 HPNWSHGLNVESTQWDDVSKMEDLNQELITLHGARNLAQSMITFNPDSIAQFGKD 606
QY 515 -----VTWIKTIRSTMIINILIVVCLFLCLLVCRCTPTAPK-----KVTSTRTG 559
DB 607 LWSHIGNWIPGLGAS-IIRYVFMFLIYLL-----TSSPKILRALWKVTSAG 654

```

```

RESULT 12
ENV_EIAYC STANDARD; PRT: 859 AA.
AC P32541;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ENV POLYPROTEIN PRECURSOR (COAT POLYPROTEIN) [CONTAINS: COAT PROTEIN
DE GP90; COAT PROTEIN GP45].
GN ENV.
OS Equine infectious anemia virus (clone CL22) (EIAV).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=31675;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92292230; PubMed=1318398;
RA Perry S.T., Flaherty M.T., Kelley M.J., Clabough D.L., Tronick S.R.,
RA Coggins L., Whetter L., Lengel C.R., Fuller F.;
RT "The surface envelope protein gene region of equine infectious anemia

```

```

RT virus is not an important determinant of tropism in vitro."
J. Virol. 66:4085-4097(1992).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M87581; AAA3005.1; -
DR PIR; C41991; VCLJ22.
DR InterPro; IPR001027; EIAV GP45.
DR InterPro; IPR001361; EIAV GP90.
DR Pfam; PF01045; EIAV GP45; 1.
DR Pfam; PF00971; EIAV GP90; 1.
KW Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 859 ENV POLYPROTEIN.
FT CHAIN 23 444 COAT PROTEIN GP90.
FT CHAIN 445 859 COAT PROTEIN GP45.
FT TRANSMEM 75 93 POTENTIAL.
FT TRANSMEM 446 472 POTENTIAL.
FT TRANSMEM 617 636 POTENTIAL.
FT TRANSMEM 787 807 POTENTIAL.
FT TRANSMEM 816 835 POTENTIAL.
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 490 490 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 859 AA; 97140 MW; 23E020E80DF334FA CRC64;

```

```

Query Match 3.6%; Score 107; DB 1; Length 859;
Best Local Similarity 20.7%; Pred. No. 0.62;
Matches 123; Conservative 79; Mismatches 244; Indels 148; Gaps 33;

QY 53 CLGRAPCCLPAVQNNLVEVTPSPNSRFTYHVMVSGMSLRPRVNYLQDFSYQSLKFRPK 112
DB 122 CWGSPGCG--RPFQNY-----PSYETNRSHMDNNTATLLE-AYHREITFIYK 166
QY 113 GK-----TCPK-EIPKSGKNTVLV-----WEECVAN---SVVILQN 145
DB 167 SSTDSHDCQEQYCKKVNLSNDSNSVRVEDVTNTAEYWGFKWLECNQTFNFTILVPE 226
QY 146 NEFTIIX---APRGQFYHNC---GOTQSCPSAQVSPAVDSLTESLDKHKHKLQSF 199
DB 227 NEWMVINDTWTIPKG-----CNETWARVKRC-----IDILYGIHPILRCVOPPE 272
QY 200 YLWEEKEGIGIS-TPR-----PKIISPVSGPEHPMLRLTVASHIRIWSGNQTLRYK 253
DB 273 FL--VQEKGIADTSRIGNCGPTIFLGLV-EDNKGVVGRDGTACNVRRLNINRKDYGIY 330
QY 254 P-FYITDLNSILTVPLQSLCKPPYMLV---GNVIKPAQITC---ENCRLETCIDSTF 307
DB 331 PIYFTCTFTNI-----TSCNNEPIISVIMYETNOV-----QYLLCNNSNNYNCVQSF 380
QY 308 N--WQHRIILLVRREGM-----WIPVSTDRPWEASPSIHILTEI 345

```



```
Db 381 GVIGQAHLELPNPKRIRNQSFNOYNCINNKTELETWKLVKT-----SGVTPLPISSEAN 436
Qy 346 KGVNLNRSKRF-IFTLIAVIMGLIAVTATAAAGVALH--SSVQSVNFVNYWQKNSTRLWN 402
Db 437 TGLIRHKRDFGISAIVAATAAATAASATSYVALTEVKNIMEVQNHTEFVENST--LN 494
Qy 403 SQSSSIDOK---LASQINDLRQTVIMWGRDLDEHHFQL-QCDWNTSDFCITPQIYNESEH 458
Db 495 GMDLIERQIKILYAMILQTHADVOLLKKEQVVEETFNIGICIERTHVFCHT-----GH 547
Qy 459 HWDVMYRRHL--QGRDNLTLDISKLKEQIFEASKAHLNLPVGTETRAIAGVADGLANLNP-- 514
Db 548 PNWMSGHLNVESTQWDDWVSMEDLNQELTLTGARNNLAQSMITFNTPDSIAQFGKDL 607
Qy 515 ---VTWIKTRSTMIINILIVVCLFLLVCRCTPTAPK-----KVTVSTRTG 559
Db 608 WSHGNWIPGLGAS-IIRYIVFMFLIYLL-----TSSPKILRALWKVTSAG 654

RESULT 13
ENV_EIAVW
ID ENV_EIAVW STANDARD; PRT; 859 AA.
AC P16082;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ENV POLYPROTEIN PRECURSOR (COAT POLYPROTEIN) [CONTAINS: COAT PROTEIN
DE GP50; COAT PROTEIN GP45].
GN ENV.
OS Equine infectious anemia virus (strain WSU5) (EIAV).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11671;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90174929; PubMed=2155398;
RA McGuire T.C., Lacy P.A., O'Rourke K.;
RT "cDNA sequence of the env gene of a pathogenic equine infectious
RT anemia lentivirus variant.";
RL Nucleic Acids Res. 18:196-196(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16988; CAA34856.1; -.
DR PIR; S07589; VCLJWS.
DR InterPro; IPR001027; EIAV_GP45.
DR InterPro; IPR001361; EIAV_GP90.
DR Pfam; PF01045; EIAV_GP45; 1.
DR Pfam; PF00971; EIAV_GP90; 1.
KW Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 859 ENV POLYPROTEIN.
FT CHAIN 23 444 COAT PROTEIN GP90.
FT CHAIN 445 859 COAT PROTEIN GP45.
FT TRANSMEM 75 93 POTENTIAL.
FT TRANSMEM 446 462 POTENTIAL.
FT TRANSMEM 614 636 POTENTIAL.
FT TRANSMEM 787 807 POTENTIAL.
FT TRANSMEM 816 835 POTENTIAL.
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 490 490 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 859 AA; 97106 MW; 6A7DBBC57050348D CRC64;

Query Match 3.5%; Score 105; DB 1; Length 859;
Best Local Similarity 20.68; Pred. No. 0.91;
Matches 123; Conservative 78; Mismatches 242; Indels 154; Gaps 33;

Qy 53 CLGRAPGCLMPAVQNLVEVPTVSPNSRFTYHVMYSGMSLRPRVNYLQDFSQVRSKLFKRPK 112
Db 122 CWGSPFGC--RPFQNY-----FSYETNRSMHMDNNTATLLE-AVHREITFIYK 166
Qy 113 GK-----TCPK-EIPKSKNTEVLV-----WEECVAN---SVVILQN 145
Db 167 SSCDSDHCQEQYCKKVNLDSSNSVRVEDVTWTAEYWGFKWLECNQOTENFKTILVPE 226
Qy 146 NEFGTIIDX---APRGFYHNCS---GOTQSCPSAQVSPAVSDSLTESLDKHKHKKLQSF 199
Db 227 NEMVINDTDWIPKG-----CNETWARVKRCP-----IDILYGIHPLRCVQPPF 272
Qy 200 YLWEEKEGIS--TPRPKIISPVSGPEHPELMRLTVASHIRIWSNQLETPRYRPFYTI 258
Db 273 FL--VQEKGIADTSR-----IGNCGPTIF-LGVLEDNKGVVGRDGTACNVSRNLNIRK 322
Qy 259 DLNSILTVP-----LQSLCKPPYMLV---GNIVIKPASQITC--ENCRFLTCD 304
Db 323 DYTGIYQPIPYTCFTFNITSCNNEPIISVIMYETNQV-----QYLLCNNNNNNNYCV 377
Qy 305 STFNL--WQHRILLVRAREGM-----WIPVSTDRPWEASPSIHILT 342
Db 378 QSFVGIVGAHLELPNPKRIRNQSFNOYNCINNKTELETWKLVKT-----SGITPLPIS 433
Qy 343 EILKGVNLNRSKRF-IFTLIAVIMGLIAVTATAAAGVALH--SSVQSVNFVNYWQKNSTR 399
Db 434 EANTGLIRHKRDFGISAIVAATAAATAASATSYVALTEVKNIMEVQNHTEFVENST- 492
Qy 400 LWNSSSIDOK---LASQINDLRQTVIMWGRDLDEHHFQL-QCDWNTSDFCITPQIYNE 455
Db 493 -LNGMDLIERQIKILYAMILQTHADVOLLKKEQVVEETFNIGICIERTHVFCHT----- 545
Qy 456 SEHHWDMYRRHL--QGRDNLTLDISKLKEQIFEASKAHLNLPVGTETRAIAGVADGLANLN 513
Db 546 -GHPWNMSGHLNVESTQWDDWVSMEDLNQELTLTGARNNLAQSMITFNTPDSIAQFG 604
Qy 514 P-----VTWIKTRSTMIINILIVVCLFLLVCRCTPTAPK-----KVTVSTRTG 559
Db 605 KDLWSHGHNWIPGLGAS-IIRYIVFMFLIYLL-----TSSPKILRALWKVTSAG 654

RESULT 14
ENV_SIVM1
ID ENV_SIVM1 STANDARD; PRT; 882 AA.
AC P05885; Q85725; Q85726;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE
DE GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)].
GN ENV.
OS Simian immunodeficiency virus (Mm142-83 isolate) (SIV-MAC).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11733;
RN [1]
```



```

QY 17 NDSVVWPGFTDRCAPKEEGMINISIGYHYPICLGRAPGCLMPAVQNWLVPTVS 76
Db 123 NDKKVAPKP-----KPAHE-----QVEPALIPS--NWTSVIPLLT 155
QY 77 PMSRFTYHMVSCM---SLRPRVNYLODF-----SYQSLKFR----- 110
Db 156 SDFKNOYSVISRLKNPNMKP-VPYAGDIKLMAFINKFSFFHSDLQNLSDQDFEVLGLD 214
QY 111 ----PKGKTCPREIPKSKNTEVLVWECVA-NSVVILQ---NNEFGTTIDXAPRGQFVH 162
Db 215 YPGDPNGSNA--GIVKGPEDTSLLLYPDFMAIKDIVYCODKMNLLFLSLDLT-----FTE 268
QY 163 NCSGQTQSCPSAQVSPAVDSLT--ESLDKHKHKLQSFYLM-----EW--- 204
Db 269 NFDGK-----SAKKKGPLTTWENL-KSSSKKVFSPNPLYRLVAREMGYPREWROO 318
QY 205 --EEKGISTPRKII-----SPVSGPEHPPELWRLTVASHHRIWNGNQTLETRYKPPYT 257
Db 319 LPSDQDISKPKTALFEQDEQTPVVDPSHPE-----ILTPNIYTWNANEPLPLE-SNPLYN 372
QY 258 --IDLNSILTVPLQSLCPYPVVLVGVNIVIKPASQTITCENCRLFTCIDSTENWQHRILL 315
Db 373 REMDANGIL-----ALKP-----MDRVVL 391
QY 316 VRAREGMWIPVSTDRPWEASPSIHILTEILK 346
Db 392 LRA-----LTD--WCASHSSAIHDEIYK 412

```

Search completed: April 9, 2002, 17:07:04
Job time: 568 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:06:31 ; Search time 70.84 Seconds
(without alignments)
1158.368 Million cell updates/sec

Title: US-09-490-700-38
Perfect score: 3011
Sequence: 1 MVTPTWMDNPIEVVNDV.....VCRCTPTAPKKTTSRTGHE 561

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_17:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2898.5	96.3	560	12 Q9QC06	Q9qc06 human endog
2	2758.5	91.6	1361	4 Q14273	Q14273 homo sapien
3	2754.5	91.5	699	4 Q9UBU4	Q9ubu4 homo sapien
4	2754.5	91.5	699	12 Q69384	Q69384 human endog
5	2746.5	91.2	694	12 Q9YNA9	Q9yna9 human endog
6	2741.5	91.0	694	12 Q9YNA5	Q9yna5 human endog
7	2733.5	90.8	694	12 Q9YNA7	Q9yna7 human endog
8	2705	89.8	698	4 Q9UKH3	Q9ukh3 homo sapien
9	2452.5	81.5	2294	4 Q9UKH9	Q9ukh9 homo sapien
10	2028	67.4	514	12 Q69386	Q69386 human endog
11	1213	40.3	245	4 Q9HDB8	Q9hdb8 homo sapien
12	955.5	31.7	271	4 Q9NX77	Q9nx77 homo sapien
13	845	28.1	277	4 Q9UKH7	Q9ukh7 homo sapien
14	840	27.9	153	4 Q95280	Q95280 homo sapien
15	840	27.9	153	12 Q42043	Q42043 unclassified
16	811	26.9	153	4 Q9UNW3	Q9unw3 homo sapien
17	805	26.7	153	4 Q9UNW2	Q9unw2 homo sapien
18	784	26.0	153	4 Q95284	Q95284 homo sapien
19	782	26.0	153	4 Q95281	Q95281 homo sapien

ALIGNMENTS

RESULT 1

Q9QC06

ID Q9QC06 PRELIMINARY; PRT; 560 AA.

AC Q9QC06;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE ENV PROTEIN.

GN ENV.

OS Human endogenous retrovirus K.

OC Viruses; Retroid viruses; Retroviridae.

OX NCBI_TaxID=45617;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=9445825; PubMed=10516026;

RA Teenjes R.R., Clauderia F., Kurth R.;

RT *Genome wide screening, cloning, chromosomal assignment and expression

RT of full-length human endogenous retrovirus type K (HERV-K)."

RL J. Virol. 73:9187-9195(1999).

DR EMBL: Y18890; CAB56604.1: -;

SQ SEQUENCE 560 AA; 63671 MW; 8A4565663901BC3A CRC64;

O95282 homo sapien
O95283 homo sapien
O9hdc0 homo sapien
O71037 human endog
O71072 human endog
O83394 mouse endog
O91zu0 endogenous
O90187 mouse mamma
O91zt5 exogenous m
O85646 mouse mamma
O91zt0 exogenous m
O991k9 mus musculu
Q9tttd3 oviss aries
Q9nlr7 oviss aries
O61536 mus musculu
Q9w121 ovine enzoo
O85643 mouse mamma
O9y135 sheep pulmo
O9yna0 sheep pulmo
O9yn97 sheep pulmo
O9w1r2 sheep pulmo
O66783 endogenous
Q9h7t6 homo sapien
Q15804 homo sapien
Q15805 homo sapien
Q9gmp3 macaca fasc

Query Match 96.3%; Score 2898.5; DB 12; Length 560;
Best Local Similarity 99.1%; Pred. No. 6.3e-261;
Matches 542; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 1 MVTPTWMDNPIEVVNDVSVVPGTDDRCAPKEEGMMINISIGYHPPICLGRPGC 60

Db 1 MVTPTWMDNPIEVVNDVSVVPGTDDRCAPKEEGMMINISIGYHPPICLGRPGC 60

Qy 61 LMPAQVQNLVVEVPVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLAFKPKGKTCPEI 120

Db 61 LMPAQVQNLVVEVPVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLAFKPKGKTCPEI 120

Qy 121 PKGSKNTFVLWEECVANSVVLQNNFETIIDAPRQGFYHNCSGQTQSCPSAQVSPAV 180

Db 121 PKGSKNTFVLWEECVANSVVLQNNFETIIDAPRQGFYHNCSGQTQSCPSAQVSPAV 180

Qy 181 DSDLTESLDKHKHKKLQSFYLWEEKEGISTPRPKIISVSGPEHPELWRLTVAHHIRI 240

Db 181 DSDLTESLDKHKHKKLQSFYLWEEKEGISTPRPKIISVSGPEHPELWRLTVAHHIRI 240

```
QY 241 WSGNOTLETRYKPYTIDLSILTVPLQSCPKPPYMLVGNIVKIPASQITTCENCRLF 300
|||||
DB 241 WSGNOTLETRYKPYTIDLSILTVPLQSCPKPPYMLVGNIVKIPASQITTCENCRLF 300
|||||
QY 301 TCIDSTFNWQHRIILVRAREGMWIPVSTDRPWEASPSHIILTEILKGVNLSKRIFITLI 360
|||||
DB 301 TCIDSTFNWQHRIILVRAREGMWIPVSTDRPWEASPSHIILTEILKGVNLSKRIFITLI 360
|||||
QY 361 AVINGLIAVATAAAGVALHSSVQSVNFVWKNSTRLWNSOSSIDOKLASQINDLRQ 420
|||||
DB 361 AVINGLIAVATAAAGVALHSSVQSVNFVWKNSTRLWNSOSSIDOKLASQINDLRQ 420
|||||
QY 421 TVIWMGDRLL-DLEHHFQLQCDWNTSDFCITPQIYNESHHWDMVRRHLQGRDNLTLDIS 479
|||||
DB 421 TVIWMGDRLL-DLEHHFQLQCDWNTSDFCITPQIYNESHHWDMVRRHLQGRDNLTLDIS 480
|||||
QY 480 KLKEQIFEASKAHLNLPVGTETAIAGVADGLANLPVTVWIKTIRSTMIINLILIVVCLFCL 539
|||||
DB 481 KLKEQIFEASKAHLNLPVGTETAIAGVADGLANLPVTVWIKTIRSTMIINLILIVVCLFCL 540
|||||
QY 540 LLVCRCT 546
|||||
DB 541 LLVCRCT 547
|||||

RESULT 2
ID Q14273 PRELIMINARY; PRT; 1361 AA.
AC Q14273;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE POL/ENV ORF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=87036922; PubMed=3021993;
RA Ono M., Tasunaga T., Miyata T., Ushikubo H.;
RT "Nucleotide sequence of human endogenous retrovirus genome related to
the mouse mammary tumor virus genome.";
RL J. Virol. 60:589-598(1986).
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
TRANSCRIPTASE)
CC EMBL; M14123; AAA8033.1; -.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_zn.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF02022; Integrase_zn; 1.
DR Pfam; PF00075; rnaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1361 AA; 153797 MW; CEB91B3F407B9498 CRC64;

Query Match 91.6%; Score 2758.5; DB 4; Length 1361;
Best Local Similarity 94.1%; Pred. No. 2.2e-247;
Matches 515; Conservative 10; Mismatches 21; Indels 1; Gaps 1;

QY 1 MVTPTWMDNPIEVYVDSVWVPGTDDRCFPAKEEGMINISIGYHYPPICLGRAPGC 60
|||||
DB 774 MVTPTWMDNPIEVYVDSVWVPGTDDRCFPAKEEGMINISIGYHYPPICLGRAPGC 833
|||||
QY 61 LMPAVQNLVVEVTPVSPNSRTYTHMVSQMSLRPRVNYLQDSYQSLKFRPKGKCPKEI 120
|||||
```

```
DB 834 LMPAVQNLVVEVTPVSPISRTYTHMVSQMSLRPRVNYLQDSYQSLKFRPKGKCPKEI 893
|||||
QY 121 PKGSKNTEVLWEECVANSVILQNNFETIIDXAPRQGFYHNCSGQTQSCPSAQVSPAV 180
|||||
DB 894 PKGSKNTEVLWEECVANSVILQNNFETIIDXAPRQGFYHNCSGQTQSCPSAQVSPAV 953
|||||
QY 181 DSDLTESLDKHKHKKLQSFYLWEEKGISTPRKIIISPVSGPEHPELWRLTVASHHRI 240
|||||
DB 954 DSDLTESLDKHKHKKLQSFYFWEWGEKISTPRKIVSPVSGPEHPELWRLTVASHHRI 1013
|||||
QY 241 WSGNOTLETRYKPYTIDLSILTVPLQSCPKPPYMLVGNIVKIPASQITTCENCRLF 300
|||||
DB 1014 WSGNOTLETRYKPYTIDLSILTVPLQSCPKPPYMLVGNIVKIPASQITTCENCRLF 1073
|||||
QY 301 TCIDSTFNWQHRIILVRAREGMWIPVSTDRPWEASPSHIILTEILKGVNLSKRIFITLI 360
|||||
DB 1074 TCIDSTFNWQHRIILVRAREGMWIPVSTDRPWEASPSHIILTEILKGVNLSKRIFITLI 1133
|||||
QY 361 AVINGLIAVATAAAGVALHSSVQSVNFVWKNSTRLWNSOSSIDOKLASQINDLRQ 420
|||||
DB 1134 AVINGLIAVATAAAGVALHSSVQSVNFVWKNSTRLWNSOSSIDOKLASQINDLRQ 1193
|||||
QY 421 TVIWMGDRLL-DLEHHFQLQCDWNTSDFCITPQIYNESHHWDMVRRHLQGRDNLTLDIS 479
|||||
DB 1194 TVIWMGDRLL-DLEHHFQLQCDWNTSDFCITPQIYNESHHWDMVRRHLQGRDNLTLDIS 1253
|||||
QY 480 KLKEQIFEASKAHLNLPVGTETAIAGVADGLANLPVTVWIKTIRSTMIINLILIVVCLFCL 539
|||||
DB 1254 KLKEQIFEASKAHLNLPVGTETAIAGVADGLANLPVTVWIKTIRSTMIINLILIVVCLFCL 1313
|||||
QY 540 LLVCRCT 546
|||||
DB 1314 LLVCRCT 1320
|||||

RESULT 3
Q9UBU4
ID Q9UBU4 PRELIMINARY; PRT; 699 AA.
AC Q9UBU4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ENVELOPE PROTEIN.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99400989; PubMed=10469592;
RA Barbulescu M., Turner G., Seaman M.I., Deinard A.S., Kidd K.K.,
RA Lenz J.;
RT "Many human endogenous retrovirus K (HERV-K) proviruses are unique to
humans.";
RL Curr. Biol. 9:861-868(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99178257; PubMed=10080172;
RA Mayer J., Sauter M., Racz A., Scherer D., Mueller-Lantzsch N.,
RA Meese E.;
RT "An almost-intact human endogenous retrovirus K on human chromosome
7.";
RL Nat. Genet. 21:257-258(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Reus K., Mayer J., Sauter M., Scherer D., Mueller-Lantzsch N.,
RA Meese E.;
RT "Further characterization of the almost intact human endogenous
retrovirus K on human chromosome 7.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF164614; AAC51798.1; -.
DR EMBL; AF074086; AAF88168.1; -.

```

DR EMBL; AF074086; AAD21098.1; -;
 KW Envelope protein.
 SQ SEQUENCE 699 AA; 79217 MW; 3844DA6D80D76DFC CRC64;

Query Match 91.5%; Score 2754.5; DB 4; Length 699;
 Best Local Similarity 94.0%; Pred. No. 2.1e-247;
 Matches 514; Conservative 10; Mismatches 22; Indels 1; Gaps 1;

Qy 1 MVTPTVMDNPIEVYVNDVSVVWPGPTDDRCAPAKPEEGMMINISIGYHYPPICLGRAPGC 60
 Db 112 LIRAVTWMNDNTEYVYVNDVSVVWPGPIDDDRCAPAKPEEGMMINISIGYHYPPICLGRAPGC 171
 Qy 61 LMPAVQNLVVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120
 Db 172 LMPAVQNLVVEVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 231
 Qy 121 PKGSKNTEVLVWEECVANSVILQNNFETIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 180
 Db 232 PKSKNTEVLVWEECVANSVILQNNFETIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 291
 Qy 181 DSDLTESLDKHKHKKLQSFYLWEEKEGISTPRPKIISPVSGPEHPELWRLTVASHHTRI 240
 Db 292 DSDLTESLDKHKHKKLQSFYPMWEGEKISTPRPKIISPVSGPEHPELWRLTVASHHTRI 351
 Qy 241 WSGNOTLETRYRKPFYITDLNSILTVPLQSCVKPPYMLVGVNIVIKPASOTITCENCRLF 300
 Db 352 WSGNOTLETRYRKPFYITDLNSILTVPLQSCVKPPYMLVGVNIVIKPASOTITCENCRLF 411
 Qy 301 TCIDSTFNWQHRIILVLRAREGWIPVSTRPWEASPSHIITLILKGVNLSKRFIFTLI 360
 Db 412 TCIDSTFNWQHRIILVLRAREGWIPVSTRPWEASPSHIITLILKGVNLSKRFIFTLI 471
 Qy 361 AVIMGLIATATAAAGVALHSSVQSVNFVWQKNSRLNWSOSSIDOKLANQINDLRQ 420
 Db 472 AVIMGLIATATAAAGVALHSSVQSVNFVWQKNSRLNWSOSSIDOKLANQINDLRQ 531
 Qy 421 TVIMWGDRL-DLEHHFQLOCDWNTSDFCITPOIYNESEHHWDMVRRHLLQREDNLTLDIS 479
 Db 532 TVIMWGDRLMSLEHFRFQLOCDWNTSDFCITPOIYNESEHHWDMVRRHLLQREDNLTLDIS 591
 Qy 480 KLKEQIFEASKAHLNLPVGTETAGVADGLANLNPVTWIKTIRSTMIINLILVVCFLCL 539
 Db 592 KLKEQIFEASKAHLNLPVGTETAGVADGLANLNPVTWIKTIRSTMIINLILVVCFLCL 651
 Qy 540 LLVCRCT 546
 Db 652 LLVCRCT 658

RESULT 4
 Q69384 ID Q69384 PRELIMINARY; PRT; 699 AA.
 AC Q69384;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE ENV MRNA.
 GN ENV.
 OS Human endogenous retrovirus.
 OC Viruses; Retrovird viruses; Retroviridae.
 OX NCBI_TaxID=11827;
 RN [1]
 RP MEDLINE=95074858; PubMed=7983704;
 RA Loewer R., Toenjes R.R., Korbmacher C., Kurth R., Loewer J.;
 RT "Identification of a Rev-related protein by analysis of spliced
 transcripts of the human endogenous retroviruses HTDV/HERV-K";
 RL J. Virol. 69:141-149(1995).
 DR EMBL; X82272; CAA57723.1; -;
 SQ SEQUENCE 699 AA; 79217 MW; 3844DA6D80D76DFC CRC64;

Query Match 91.5%; Score 2754.5; DB 12; Length 699;
 Best Local Similarity 94.0%; Pred. No. 2.1e-247;
 Matches 514; Conservative 10; Mismatches 22; Indels 1; Gaps 1;

Qy 1 MVTPTVMDNPIEVYVNDVSVVWPGPTDDRCAPAKPEEGMMINISIGYHYPPICLGRAPGC 60
 Db 112 LIRAVTWMNDNTEYVYVNDVSVVWPGPIDDDRCAPAKPEEGMMINISIGYHYPPICLGRAPGC 171
 Qy 61 LMPAVQNLVVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 120
 Db 172 LMPAVQNLVVEVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYQSLKFRPKGKTCPCKEI 231
 Qy 121 PKGSKNTEVLVWEECVANSVILQNNFETIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 180
 Db 232 PKSKNTEVLVWEECVANSVILQNNFETIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 291
 Qy 181 DSDLTESLDKHKHKKLQSFYLWEEKEGISTPRPKIISPVSGPEHPELWRLTVASHHTRI 240
 Db 292 DSDLTESLDKHKHKKLQSFYPMWEGEKISTPRPKIISPVSGPEHPELWRLTVASHHTRI 351
 Qy 241 WSGNOTLETRYRKPFYITDLNSILTVPLQSCVKPPYMLVGVNIVIKPASOTITCENCRLF 300
 Db 352 WSGNOTLETRYRKPFYITDLNSILTVPLQSCVKPPYMLVGVNIVIKPASOTITCENCRLF 411
 Qy 301 TCIDSTFNWQHRIILVLRAREGWIPVSTRPWEASPSHIITLILKGVNLSKRFIFTLI 360
 Db 412 TCIDSTFNWQHRIILVLRAREGWIPVSTRPWEASPSHIITLILKGVNLSKRFIFTLI 471
 Qy 361 AVIMGLIATATAAAGVALHSSVQSVNFVWQKNSRLNWSOSSIDOKLANQINDLRQ 420
 Db 472 AVIMGLIATATAAAGVALHSSVQSVNFVWQKNSRLNWSOSSIDOKLANQINDLRQ 531
 Qy 421 TVIMWGDRL-DLEHHFQLOCDWNTSDFCITPOIYNESEHHWDMVRRHLLQREDNLTLDIS 479
 Db 532 TVIMWGDRLMSLEHFRFQLOCDWNTSDFCITPOIYNESEHHWDMVRRHLLQREDNLTLDIS 591
 Qy 480 KLKEQIFEASKAHLNLPVGTETAGVADGLANLNPVTWIKTIRSTMIINLILVVCFLCL 539
 Db 592 KLKEQIFEASKAHLNLPVGTETAGVADGLANLNPVTWIKTIRSTMIINLILVVCFLCL 651
 Qy 540 LLVCRCT 546
 Db 652 LLVCRCT 658

RESULT 5
 Q9YNA9 ID Q9YNA9 PRELIMINARY; PRT; 694 AA.
 AC Q9YNA9;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE ENV PROTEIN.
 GN ENV.
 OS Human endogenous retrovirus K.
 OC Viruses; Retrovird viruses; Retroviridae.
 OX NCBI_TaxID=45617;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Toenjes R.R., Czauderna F., Kurth R.;
 RT "Full-length human endogenous retrovirus type K (HERV-K) elements
 encoding Gag, Pol and Env proteins are localised on chromosomes 7 and
 19.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Y17832; CAA76880.1; -;
 SQ SEQUENCE 694 AA; 78690 MW; 82P91825669CF25B CRC64;

Query Match 91.2%; Score 2746.5; DB 12; Length 694;
 Best Local Similarity 93.8%; Pred. No. 1.2e-246;
 Matches 513; Conservative 10; Mismatches 23; Indels 1; Gaps 1;

Qy 1 MVTPTVMDNPIEVYVNDVSVVWPGPTDDRCAPAKPEEGMMINISIGYHYPPICLGRAPGC 60

```
Db 107 LIRAVTMDNTEVYVNDVSVVWPGIDDRCPAKPPEEGMMINISIGYHPPICLGRAPGC 166
QY 61 LMPAVONMLVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSRSLKFRPKGKTCPKREI 120
Db 167 LMPAVONMLVEVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYQSRSLKFRPKGKTCPKREI 226
QY 121 PKGSKNTEVLWEECVANSVILQNNFETIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 180
Db 227 PKSKNTEVLWEECVANSVILQNNFETIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 286
QY 181 DSDLTSLDKHKHKKLQSFYLMWEEKGIISTPRKIIISPVSGPEHPPELWRLTVASHHRI 240
Db 287 DSDLTSLDKHKHKKLQSFYLMWEEKGIISTPRKIIISPVSGPEHPPELWRLTVASHHRI 346
QY 241 WSGNOTLETTRYKPKFYITIDLSILVPLQSCVLPYMLVGVNIVKPSQITCENCRL 300
Db 347 WSGNOTLETTRYKPKFYITIDLSILVPLQSCVLPYMLVGVNIVKPSQITCENCRL 406
QY 301 TCIDSTFNWQHRIILVAREGWIIPVSDRPWEASPSIHLTEILKGVNLSKRFTFTLI 360
Db 407 TCIDSTFNWQHRIILVAREGWIIPVSDRPWEASPSIHLTEILKGVNLSKRFTFTLI 466
QY 361 AVINGLIAVATAAAGVALHSSVQSVNFVWQKNSTRNLWNSQSSIDOKLASQINDLRO 420
Db 467 AVINGLIAVATAAAGVALHSSVQSVNFVWQKNSTRNLWNSQSSIDOKLASQINDLRO 526
QY 421 TVIWMGDLRL-DEHFFQLQCDWNTSDFCITPOIYNESEHHWDMVRRHLOGREDNLTLDIS 479
Db 527 TVIWMGDLRL-DEHFFQLQCDWNTSDFCITPOIYNESEHHWDMVRRHLOGREDNLTLDIS 586
QY 480 KLKEQIFEASKAHLNLPVGTAEIAGVADGLANLNPVTWIKTIRSTMIINLILVVCFLCL 539
Db 587 KLKEQIFEASKAHLNLPVGTAEIAGVADGLANLNPVTWIKTIRSTMIINLILVVCFLCL 646
QY 540 LLVCRCT 546
Db 647 LLVCRCT 653
```

```
RESULT 6
QYNA5 ID QYNA5 PRELIMINARY: PRT; 694 AA.
AC QYNA5;
DT 01-MAY-1999 (TremBLrel. 10, Created)
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TremBLrel. 10, Last annotation update)
DE ENV PROTEIN.
GN ENV.
OS Human endogenous retrovirus K.
OC Viruses; Retroviral viruses; Retroviridae.
OX NCBI_TaxID=45617;
RN [1]
RP SEQUENCE FROM N.A.
RA Toenjes R.R., Czauderna F., Kurth R.;
RT "Full-length human endogenous retrovirus type K (HERV-K) elements encoding Gag, Pol and Env proteins are localised on chromosomes 7 and 19."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y17834; CAA76886.1; -.
SQ SEQUENCE 694 AA; 78635 MW; 23A8A5A188543317 CRC64;
```

```
Query Match 91.0%; Score 2741.5; DB 12; Length 694;
Best Local Similarity 93.6%; Pred. No. 3.4e-246;
Matches 512; Conservative 11; Mismatches 23; Indels 1; Gaps 1;

QY 1 MVTPTWMDNPIEVYVNDVSVVWPGIDDRCPAKPPEEGMMINISIGYHPPICLGRAPGC 60
Db 107 LIRAVTMDNTEVYVNDVSVVWPGIDDRCPAKPPEEGMMINISIGYHPPICLGRAPGC 166
QY 61 LMPAVONMLVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSRSLKFRPKGKTCPKREI 120
Db 167 LMPAVONMLVEVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYQSRSLKFRPKGKTCPKREI 226
QY 121 PKGSKNTEVLWEECVANSVILQNNFETIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 180
Db 227 PKSKNTEVLWEECVANSVILQNNFETIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 286
```

```
Db 167 IMPAVONMLVEVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYQSRSLKFRPKGKTCPKREI 226
QY 121 PKGSKNTEVLWEECVANSVILQNNFETIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 180
Db 227 PKSKNTEVLWEECVANSVILQNNFETIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 286
QY 181 DSDLTSLDKHKHKKLQSFYLMWEEKGIISTPRKIIISPVSGPEHPPELWRLTVASHHRI 240
Db 287 DSDLTSLDKHKHKKLQSFYLMWEEKGIISTPRKIIISPVSGPEHPPELWRLTVASHHRI 346
QY 241 WSGNOTLETTRYKPKFYITIDLSILVPLQSCVLPYMLVGVNIVKPSQITCENCRL 300
Db 347 WSGNOTLETTRYKPKFYITIDLSILVPLQSCVLPYMLVGVNIVKPSQITCENCRL 406
QY 301 TCIDSTFNWQHRIILVAREGWIIPVSDRPWEASPSIHLTEILKGVNLSKRFTFTLI 360
Db 407 TCIDSTFNWQHRIILVAREGWIIPVSDRPWEASPSIHLTEILKGVNLSKRFTFTLI 466
QY 361 AVINGLIAVATAAAGVALHSSVQSVNFVWQKNSTRNLWNSQSSIDOKLASQINDLRO 420
Db 467 AVINGLIAVATAAAGVALHSSVQSVNFVWQKNSTRNLWNSQSSIDOKLASQINDLRO 526
QY 421 TVIWMGDLRL-DEHFFQLQCDWNTSDFCITPOIYNESEHHWDMVRRHLOGREDNLTLDIS 479
Db 527 TVIWMGDLRL-DEHFFQLQCDWNTSDFCITPOIYNESEHHWDMVRRHLOGREDNLTLDIS 586
QY 480 KLKEQIFEASKAHLNLPVGTAEIAGVADGLANLNPVTWIKTIRSTMIINLILVVCFLCL 539
Db 587 KLKEQIFEASKAHLNLPVGTAEIAGVADGLANLNPVTWIKTIRSTMIINLILVVCFLCL 646
QY 540 LLVCRCT 546
Db 647 LLVCRCT 653
```

```
RESULT 7
QYNA7 ID QYNA7 PRELIMINARY: PRT; 694 AA.
AC QYNA7;
DT 01-MAY-1999 (TremBLrel. 10, Created)
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)
DE ENV PROTEIN.
GN ENV.
OS Human endogenous retrovirus K.
OC Viruses; Retroviral viruses; Retroviridae.
OX NCBI_TaxID=45617;
RN [1]
RP SEQUENCE FROM N.A.
RA Toenjes R.R., Czauderna F., Kurth R.;
RT "Full-length human endogenous retrovirus type K (HERV-K) elements encoding Gag, Pol and Env proteins are localised on chromosomes 7 and 19."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y17833; CAA76883.1; -.
SQ SEQUENCE 694 AA; 78692 MW; 498A1D74536D0EE CRC64;
```

```
Query Match 90.8%; Score 2733.5; DB 12; Length 694;
Best Local Similarity 93.4%; Pred. No. 1.9e-245;
Matches 511; Conservative 10; Mismatches 25; Indels 1; Gaps 1;

QY 1 MVTPTWMDNPIEVYVNDVSVVWPGIDDRCPAKPPEEGMMINISIGYHPPICLGRAPGC 60
Db 107 LIRAVTMDNPIEVYVNDVSVVWPGIDDRCPAKPPEEGMMINISIGYHPPICLGRAPGC 166
QY 61 LMPAVONMLVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQSRSLKFRPKGKTCPKREI 120
Db 167 LMPAVONMLVEVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYQSRSLKFRPKGKTCPKREI 226
QY 121 PKGSKNTEVLWEECVANSVILQNNFETIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 180
Db 227 PKSKNTEVLWEECVANSVILQNNFETIIDXAPRGQFYHNCSGQTQSCPSAQVSPAV 286
```



```
QY 61 LMPAVQNLVEVPTVSPNSRFTYHVMGSLRPRVNYLQDFSYQSRSLKFRPKGKCPKEI 120
|||||
DB 61 LMPAVQNLVEVPTVSPNSRFTYHVMGSLRPRVNYLQDFSYQSRSLKFRPKGKCPKEI 120
|||||
QY 121 PKGSKNTEVLWEECVANSVVLQNNFEGTIDXAPRGQYHNCSGGTQSCPSAQVSPAV 180
|||||
DB 121 PKGSKNTEVLWEECVANSVVLQNNFEGTIDXAPRGQYHNCSGGTQSCPSAQVSPAV 180
|||||
QY 181 DSDLTSLDKHKKHKLQSFYLWEEKGISTPRPKIISPVSGPEHPELWRL 231
|||||
DB 181 DSDLTSLDKHKKHKLQSFYEWEGKISTPRPEIISPVSGPEHPELWRL 231
|||||
RESULT 12
Q9NX77
ID Q9NX77 PRELIMINARY; PRT; 271 AA.
AC Q9NX77;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CDNA FLJ20393 FIS, CLONE KAITA4699.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ILEAL MUCOSA;
RA Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K.,
RA Hirao M., Ohmori Y., Ota T., Suzuki Y., Oabayashi M., Nishi T.,
RA Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000400; BRA91140.1; -.
SQ SEQUENCE 271 AA; 30693 MW; EC1590F3D599660E CRC64;
```

```
Query Match 31.7%; Score 955.5; DB 4; Length 271;
Best Local Similarity 64.4%; Pred. No. 1.3e-80;
Matches 174; Conservative 46; Mismatches 49; Indels 1; Gaps 1;

QY 277 MLVVGNTVIKPSQTITCNCRLFTCIDSTFNMOHRLILVRAREGMWIPVSTDRPWEASP 336
|||||
DB 1 MLVVGNTIIMPDQAQITIECHNCKLFTCIDATFNPTTSLILVRAREGMWIPVSLHRPWESSP 60
|||||
QY 337 SIHLTLEILKGLVLRNKRFTLTLAVTMGLIATATATAAGVALHSSVQSVNFVNYWQKN 396
|||||
DB 61 SIHLVNEVLKDLKTRKRTFTLILAVLAGLLAVTATAATAGVAIRSSVQTAHYVEACOKN 120
|||||
QY 397 STRLWNSQSSIDQKLASQINDLRQTVIWMGDR-LDLEHFFQLQCDWNTSDFCITPQIYNE 455
|||||
DB 121 SSRLWNSQAQIDQLANQINDLRQSVTWLGDVNVNLOHRLQMLQCDWNTSDYCIPTPAYNQ 180
|||||
QY 456 SEHWDVRRHLOQRENDLTLDSKLEQIFPEASKAHLNLVPGTEATAGVADGLANLNPV 515
|||||
DB 181 DQHSWENVSRLKAWDNDLTLDSIQLEQIFPEASQAHLSTVPGSHIFEGITKQLPDENPF 240
|||||
QY 516 TWIKTIRSTMIINILIVLVCFLLLVCR 545
|||||
DB 241 KWLKPVRGSLLLALLLIVLVCFLLLVCR 270
|||||
```

```
RESULT 13
Q9UKH7
ID Q9UKH7 PRELIMINARY; PRT; 277 AA.
AC Q9UKH7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE ENVELOPE PROTEIN.
GN ENV.
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99400989; PubMed=10469592;
RA Barbolescu M., Turner G., Seaman M.I., Deinard A.S., Kidd K.K.,
RA Lenz J.;
RT "Many human endogenous retrovirus K (HERV-K) proviruses are unique to
RT humans";
RT Curr. Biol. 9:861-868(1999).
DR EMBL; AF164612; AAD51795.1; -.
KW Envelope protein.
SQ SEQUENCE 277 AA; 31420 MW; E2FCF565CCFA85E5 CRC64;
```

```
Query Match 28.1%; Score 845; DB 4; Length 277;
Best Local Similarity 92.8%; Pred. No. 2.5e-70;
Matches 154; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 MVTPTVWMDNPIEVYVNDVSVVWPGTDDRCPAKPEEGMMINISIGYHYPPICLGRAPGC 60
|||||
DB 112 LIWAVTWMDNPIEVYVNDVSVVWPGTDDRCPAKPEEGMMINISIGYHYPPICLGRAPGC 171
|||||
QY 61 LMPAVQNLVEVPTVSPNSRFTYHVMGSLRPRVNYLQDFSYQSRSLKFRPKGKCPKEI 120
|||||
DB 172 LMPAVQNLVEVPTVSPISRFTYHVMGSLRPRVNYLQDFSYQSRSLKFRPKGKCPKEI 231
|||||
QY 121 PKGSKNTEVLWEECVANSVVLQNNFEGTIDXAPRGQYHNCSG 166
|||||
DB 232 PKGSKNTEVLWEECVANSVVLQNNFEGTIDWAPRGQYHNCSG 277
|||||
```

```
RESULT 14
O95280
ID O95280 PRELIMINARY; PRT; 153 AA.
AC O95280;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE ENVELOPE PROTEIN RIC-1.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449692; PubMed=9778243;
RA Lan M.S., Mason A., Coutant R., Chen Q.Y., Vargas A., Rao J.,
RA Gomez R., Chalew S., Garry R., MacLaren N.K.;
RT "HERV-K10s and immune-mediated (type 1) diabetes.";
RL Cell 95:14-16(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Hasuike S., Jinno Y.;
RT "Isolation and localization of an endogenous retrovirus gene, a
RT candidate gene for type 1 diabetes, and identification of a CA repeat
RT marker at its locus.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF084864; AAC68893.1; -.
DR EMBL; AF134984; AAD33055.1; -.
KW Envelope protein.
SQ SEQUENCE 153 AA; 17305 MW; 2B36E1656DF0C01A CRC64;
```

```
Query Match 27.9%; Score 840; DB 4; Length 153;
Best Local Similarity 100.0%; Pred. No. 3.2e-70;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVTPTVWMDNPIEVYVNDVSVVWPGTDDRCPAKPEEGMMINISIGYHYPPICLGRAPGC 60
|||||
DB 1 MVTPTVWMDNPIEVYVNDVSVVWPGTDDRCPAKPEEGMMINISIGYHYPPICLGRAPGC 60
|||||
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 16:56:43 ; Search time 72.35 Seconds

(without alignments)
574.362 Million cell updates/sec

Title: US-09-490-700-38

Perfect score: 3011

Sequence: 1 MVTPTWMDNPIEVVDSV.....VCRCTPTAKKTVTSRTGHE 561

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A_Geneseq_1101.*
- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
 - 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	3009	99.9	561 20 AAW95693	Human endogenous I
2	3009	99.9	561 20 AAW97746	Human endogenous I
3	1325.5	44.0	312 22 AAM14899	Peptide #1333 enco
4	1325.5	44.0	312 22 AAM27328	Peptide #1365 enco
5	1325.5	44.0	312 22 AAM02623	Peptide #1305 enco
6	840	27.9	153 20 AAW95692	Human endogenous I
7	840	27.9	153 20 AAW97745	Human endogenous I
8	840	27.9	181 20 AAW95694	Human endogenous I
9	840	27.9	181 20 AAW97747	Human endogenous I
10	557	18.5	141 22 AAB93186	Human protein sequ
11	398	13.2	688 20 AAY41142	Mouse mammary tumo

12	345	11.5	615	22	AAB69699	JRSV ENV protein.
13	340.5	11.3	615	14	AAR31349	Jaagsiekte retrovi
14	260	8.6	106	22	AAB95842	Human protein sequ
15	222.5	7.4	276	22	AAB95817	Human protein sequ
16	220	7.3	48	22	AAM14052	Peptide #486 enco
17	220	7.3	48	22	AAM26458	Peptide #495 enco
18	220	7.3	48	22	AAM01792	Peptide #474 enco
19	209	6.9	57	21	AAO03399	Human secreted pro
20	140.5	4.7	59	20	AAV35898	Extended human sec
21	136.5	4.5	1123	20	AAV41086	Human lectomedin-1
22	136.5	4.5	1123	21	AAB15724	Human lectomedin-1
23	135.5	4.5	1114	20	AAV41092	Peptide Seq ID No:
24	135.5	4.5	1114	21	AAB15730	Protein encoded by
25	135.5	4.5	1177	20	AAV41085	Human lectomedin-1
26	135.5	4.5	1177	21	AAB15723	Human lectomedin-1
27	135.5	4.5	1403	20	AAV41087	Human lectomedin-1
28	135.5	4.5	1403	21	AAB15725	Human lectomedin-1
29	115.5	3.8	153	22	AAB67118	Rhesus mammary tum
30	112.5	3.7	152	22	AAB67127	Human mammary tumo
31	112.5	3.7	153	22	AAB67115	Human mammary tumo
32	112.5	3.7	153	22	AAB67125	Human mammary tumo
33	112	3.7	100	21	AAB44393	Sequence homologou
34	111.5	3.7	1466	20	AAV41091	Rat latrophilin pr
35	111.5	3.7	1466	21	AAB15729	Rat latrophilin
36	111.5	3.7	1471	19	AAW74585	Calcium independen
37	110.5	3.7	153	22	AAB67114	Mouse mammary tumo
38	110.5	3.7	153	22	AAB67117	Human mammary tumo
39	110.5	3.7	153	22	AAB67126	Human mammary tumo
40	108.5	3.6	153	22	AAB67119	Rhesus mammary tum
41	108.5	3.6	2289	18	AAW14987	Protein derived fr
42	107.5	3.6	153	22	AAB67120	Cat mammary tumour
43	107.5	3.6	153	22	AAB67121	Cat mammary tumour
44	105	3.5	735	9	AAW81758	Sequence encoded b
45	104.5	3.5	863	21	AAB35756	EIAV env gene prod

ALIGNMENTS

RESULT 1
AAW95693
ID AAW95693 standard; Protein: 561 AA.
XX AC AAW95693;
XX AC AAW95693;
DT 08-JUN-1999 (first entry)
XX Human endogenous retrovirus IDDMK1.2-22 env/fs-sag protein.
DE IDDM; insulin-dependent diabetes mellitus; endogenous retrovirus;
KW SAG; superantigen; provirus; autoimmune disease; type 1 diabetes;
KW diagnosis; env; envelope.
XX Homo sapiens.
OS Homo sapiens.
XX EP893691-A1.
PN EP893691-A1.
XX 27-JAN-1999.
PD 27-JAN-1999.
XX 23-JUL-1997; 97EP-0401773.
PF 23-JUL-1997; 97EP-0401773.
XX 23-JUL-1997; 97EP-0401773.
PR (MACH/) MACH B F.
XX Conrad B, Mach B;
PI WPI; 1999-097928/09.
DR N-PSDB; AAX07515.
XX Diagnosing human autoimmune disease by detecting retrovirus with
PT superantigen activity - new retrovirus associated with type 1
PT diabetes, its proviral DNA, and related vectors, transformed cells,

PT proteins, antibodies and specific binding agents, used for treating
XX or preventing autoimmune disease
PS Claim 31; Fig 7E; 92pp; English.
XX
CC The sequence is that of an insulin-dependent diabetes mellitus
CC associated human endogenous retrovirus (IDDMK1.2-22) env/fs-sag protein.
CC The retrovirus has Superantigen (Sag) activity. It can be used
CC as part of a method is specifically used to diagnose type 1 diabetes
CC mellitus. Modified proteins expressed by the retroviral sequence
CC (without SAG activity but still able to induce an immune response)
CC are useful in vaccines to treat or prevent SAG-related autoimmune
CC disease; nucleic acid sequences encoding (modified) SAG can be used
CC similarly to treat such diseases. Retroviral-encoded SAG are important
CC in pathogenesis of autoimmune disease. Retroviral-activated
CC autoreactive T cells. The method is very specific (it can differentiate
CC between expressed and non-expressed viral nucleic acids) and can be used
CC even where the pathogen is an ubiquitous endogenous retrovirus. Blood
CC or plasma samples can be tested without extensive preparation and
CC diagnosis can be made before clinical signs are apparent, allowing
CC early intervention before severe tissue damage has occurred.
XX
SQ Sequence 561 AA;

Query Match 99.9%; Score 3009; DB 20; Length 561;
Best Local Similarity 99.8%; Pred. No. 1e-283;
Matches 560; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVTPTWMDNPIEVYVNDVSVVPGTDDRCAPKPEEGGMINISIGYHYPPICLGRAPGC 60
Db 1 mvtptwmdnplievynvndsvvpgtddrcapkppeeggmminisigyhyppiclgrapgc 60
QY 61 LMPAVONHLEVPVTPSPNSRFTYHMVSGMSLRPRVNYLQDFSYORSKFRPKGKTCPEI 120
Db 61 lmpavonhlevptvspnsrftymvsgmslrprvnylqdfsyorskfrpkgtcpei 120
QY 121 PKGSKNTEVLWEECVANSVILQNNFEGTTPVQVQFYNHCSGQTSQSPSAQVSPAV 180
Db 121 pkgskntevlwecvansvvlqnnfegtppvqvqfynhcsqgtsqspsaqvspav 180
QY 181 DSDLTESLDKHKHKLQSFYLWEEKGISTPRKIISPVSGPEHPELWRTVASHHRI 240
Db 181 dsdltesldkhhklqsfylweeekgistprkiispsvsgpehpelwrtvashhiri 240
QY 241 WSGNQTLTRYKPFYITDLNSILTVPLQSLKPPYMLVGNIVIKPASQITCENCRLF 300
Db 241 wsgnqtlettrykpfyitdlnsiltvplqslkppymlvgnivikpasqitcencrlf 300
QY 301 TCIDSTFNWQHRIILVRAREGMWITPVSDRPAEASPSHILTEILKGVNLNRSKRIFTLI 360
Db 301 tcidstfnwqhrillvraregmwvitpvdsdrpaeaspsihlileilkgvnlnrskrifl 360
QY 361 AVINGLTAVTATAAGVALHSSVQSVNFVNYWKNSTRNLWSSOISDOKLASQINDLRQ 420
Db 361 avingltavtataagvalhssvqsvnfvnynwknstrnlwssosidoklasqindlrq 420
QY 421 TVIWMGDRDLDEHHFQLOCDWNTSDFCITPQIYNESEHHDWVRHLLQGRDNLTLDISK 480
Db 421 tviwmgdrdldehhfqlodcwnntsdfcftpqiynesehhdwvrhllqgrdnltldisk 480
QY 481 LKEQIFEASKAHLNLVPGTEAIGVAGLANLNPVWIKTIRSTMIINLILIVVCLFCLL 540
Db 481 lkeqifeaskahlnlvpgteaigvaglanlnpvtwiktirstmiinlilivvclfccll 540
QY 541 LVCRCTPTAPKKTTSRTGHE 561
Db 541 lvcrctptapkktvtsrtghe 561

RESULT 2
AAW97746
ID AAW97746 standard; Protein: 561 AA.

XX AAW97746;
XX 21-MAY-1999 (first entry)
XX Human endogenous retrovirus IDDMK1.2-22 envelope protein.
XX HERV; IDDMK1.2-22; superantigen; SAG; antigen; IDDM;
XX insulin-dependent diabetes mellitus; autoimmune disease; diagnosis;
XX therapy; vaccine; envelope protein; env gene.
XX Human endogenous retrovirus.
XX Key Location/Qualifiers
XX Modified-site 17..19 /note= "Asn is N-glycosylated"
XX Modified-site 42..45 /note= "Asn is N-glycosylated"
XX W09905527-A2.
XX 04-FEB-1999.
XX 22-JUL-1998; 98WO-EP04926.
XX 23-JUL-1997; 97EP-0401773.
XX 22-JUL-1997; 97EP-0112482.
XX (MEDI-) MEDTGEN SA.
XX Conrad B, Mach B;
XX WPI; 1999-143118/12.
XX New isolated human endogenous retrovirus - used to develop products
XX for the diagnosis, prevention and treatment of autoimmune disease,
XX particularly insulin dependent diabetes mellitus
XX Claim 6; Fig 7E; 165pp; English.
XX This is the envelope protein (Env) of a new human endogenous
XX retrovirus (HERV), designated IDDMK1.2-22, that has been identified
XX as the source of superantigen (SAG) activity in insulin-dependent
XX diabetes mellitus (IDDM) patients. The endogenous retrovirus is
XX ubiquitous in the human genome but is only expressed in diabetic
XX individuals. The HERV encodes SAG activity within the env gene.
XX A claimed process for the diagnosis, including the pre-symptomatic
XX diagnosis, of a human autoimmune disease associated with a HERV
XX having SAG activity comprises specifically detecting in a
XX biological sample either: (a) the mRNA of an expressed HERV having
XX SAG activity (especially IDDMK1.2-22 5'LTR, 3' long terminal
XX repeat, env or pol); (b) a protein or peptide expressed by the HERV
XX (see AAW97745-48); (c) antibodies specific to the proteins expressed
XX by the HERV; or (d) SAG activity specifically associated with the
XX HERV. Products of the invention can be used to identify substances
XX capable of blocking transcription or translation of SAG-encoding
XX nucleic acid sequences, useful in therapy and/or prevention of
XX autoimmune disease associated with the SAG. A nucleic acid encoding
XX human retroviral SAG can be used as a DNA vaccine. Expression of
XX the endogenous SAG in IDDM suggests a general model according to
XX which self SAG-driven and systemic activation of autoreactive T
XX cells leads to organ-specific autoimmune disease.
XX Sequence 561 AA;

Query Match 99.9%; Score 3009; DB 20; Length 561;
Best Local Similarity 99.8%; Pred. No. 1e-283;
Matches 560; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVTPTWMDNPIEVYVNDVSVVPGTDDRCAPKPEEGGMINISIGYHYPPICLGRAPGC 60
Db 1 mvtptwmdnplievynvndsvvpgtddrcapkppeeggmminisigyhyppiclgrapgc 60

Qy 61 LMPAVQNLWVEVPTVSPNSRETYHMVSGMSLRPRVNYLQDFSYORSKFRPKGKTCPKREI 120
Db 61 Lmpavqnlwvevptvspnsrtyhmvgmslrprvnylqdfsyqrskfrpkgtcpkei 120
Qy 121 PKGSKNTEVLVWEECVANSVILQNNRGTIIDXAPRGQFYHNCSGQQTQSCPSAQVSPAV 180
Db 121 pkgskntevlvweecvansvillqnnrftiidzaprgqfynhcsqgqtcscpsaqvspav 180
Qy 181 DSDLTESLDKHKHKLQSFYLWEWEEKIGISTPRPKIISPVSQPEHPELWRLTVAASHHTRI 240
Db 181 dsdltesldkhkhklqsfylweewekigstprpkispsvgpenpelwrltvasashhtri 240
Qy 241 WSGNQLETRYKPFYTTIDLSILTPLOQCLKPPYMLVGVNIVKIPASQITTCENCRLF 300
Db 241 wsgnqletryrkpfytdlslsiltploqscldkppymlvvgnivkpasqitcencrlf 300
Qy 301 TCIDSTFNWQHRILLVRAREGMWIPVSTDRPWEASPSHIILTEILKGVNLSKRFIFTLI 360
Db 301 tcidstfnwqhrillvraregmwipvstdrpweaspsihlilteilkgvlnrskrfiftli 360
Qy 361 AVINGLIATATAAAGVALHSSVQSVNFVWQKNSTRLWNSOSSIDQKLASQINDLRQ 420
Db 361 avingliavtataaagvalhssvqsvnfvwqknstrlwnsossidqklasqindlrq 420
Qy 421 TVIWMGDRDLDEHHFQLQCDWNTSDFCITPQIYNESEHHWDMVRRHLOGREDNLTLDISK 480
Db 421 tvwmgdrldlehhfqlqcdwntsfctipqiynesehhwdmvrhlogrednltldisk 480
Qy 481 LKEQIFEASKAHLNLPVQTEAIAGVADGLANLPVTWIKTIRSTMIINLIIVVCLFCLL 540
Db 481 lkeqifeaskahlnlpvqteaiagvadglanlpvtwiktirstmiinliivvclfcll 540
Qy 541 LVCRCPTAPKKTVTSRTGHE 561
Db 541 lvcrcptapkktvtsrtghe 561

RESULT 3
AAM14899
ID AAM14899 standard; Protein; 312 AA.
XX
AC AAM14899;
XX
DT 12-OCT-2001 (first entry)
XX
DE Peptide #1333 encoded by probe for measuring cervical gene expression.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX

PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
PS Claim 27; SEQ ID NO 19725; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs: see AAI10068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 312 AA;
Query Match 44.0%; Score 1325.5; DB 22; Length 312;
Best Local Similarity 93.7%; Pred. No. 2.5e-120;
Matches 254; Conservative 6; Mismatches 10; Indels 1; Gaps 1;
Qy 277 MLVVGNIIVIKPASQITTCENCRLEFTCIDSTFNWQHRILLVRAREGMWIPVSTDRPWEASP 336
Db 1 mlvvgnivikpdsqitcencrlltcidstfnwqhrillvraregvwllvsmdrpweasp 60
Qy 337 SIHILTEILKGVNLSKRFIFTLIAVIMGLIAVTATAAAGVALHSSVQSVNFVWQKN 396
Db 61 siviltevikgvlnrskrfiftliavimgliavtatagavallhssvqsvnfvdwqkn 120
Qy 397 STRLWNSOSSIDQKLASQINDLRQTVIWMGDRDL-DEHHFQLQCDWNTSDFCITPQIYNE 455
Db 121 strlwnsossidqklasqindlrqtvwmgdrilmslehrfqlqcdwnscdfcftpqiyn 180
Qy 456 SEHHWDMVRRHLOGREDNLTLDISKLEQIFEASKAHLNLPVQTEAIAGVADGLANLPV 515
Db 181 sehwdmvrhlogrednltldiskleqifeaskahlnlpvqteaiagvadglanlpv 240
Qy 516 TWIKTIRSTMIINLIIVVCLFCLLVCRCT 546
Db 241 twiktigstmiinliivclfcillvcrct 271

RESULT 4
AAM27328
ID AAM27328 standard; Protein; 312 AA.
XX
AC AAM27328;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #1365 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX

```

XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-488897/53.
XX DR Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PT Claim 27; SEQ ID No 27597; 654pp; English.
XX PS The present invention relates to single exon nucleic acid probes (SENP:
XX CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders.
XX SQ Sequence 312 AA;

Query Match 44.0%; Score 1325.5; DB 22; Length 312;
Best Local Similarity 93.7%; Pred. No. 2.5e-120;
Matches 254; Conservative 6; Mismatches 10; Indels 1; Gaps 1;

Qy 277 MLVGVNIVIKPASOTITCENCRLFTCIDSTFNWQHRIILVRAREGMWIPVSTDRPWEASP 336
Db 1 mlvvgnivikpdsqtqtcencrlltcidstfnwqhrillvraregvwllvsmndrpweasp 60

Qy 337 SIHILTEILKGVNRSKRFIPTLIIVIMGLIATVATAAGVALHSSVQSVNFVWYWK 396
Db 61 svhiltelkvgvnrskrfiftliavimgliatvataagvalhssvqsvnfvndwqkn 120

Qy 397 STRLWNSOSSIDQKLASQINDLRQTIVWMDRL-DLEHHFQLQCDWNTSDFCITPQIYNE 455
Db 121 strlwnsqssidqkqlangindlrqtivwmdrflmslehnrfqlqcdwntsdftcpqiye 180

Qy 456 SEHHWDMVRRHLQGRDNLTLDISKLEQIFEASKAHNLNVPGETAAGVADGLANLPV 515
Db 181 sehhdwmvrrhlqgrdnltldiskleqifeaskahlnlvpgteaagvadglaanlpv 240

Qy 516 TWIKTIRSTMIINLILVVCFLLLVCRCT 546
Db 241 twktigtsttiinlililvcfllllvcrct 271

RESULT 5
AAW02623
ID AAW02623 standard; Protein; 312 AA.
XX AC AAW02623;
XX DT 09-OCT-2001 (first entry)
XX DE Peptide #1305 encoded by probe for measuring breast gene expression.
XX KW Probe; human; breast disease; breast cancer; development disorder;
XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN WO200157270-A2.
XX PD 09-AUG-2001.
XX PF 29-JAN-2001; 2001WO-US00661.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608406.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.

```

```

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-476286/51.
XX PT Novel single exon nucleic acid probe used to measuring gene expression
XX PS in a human breast -
XX PS Claim 27; SEQ ID No 11363; 322pp; English.
XX CC The present invention relates to novel single exon nucleic acid probes
XX CC (see AAI00010-AAI10067). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for measuring human gene expression in
XX CC a human breast sample, where the probe hybridises at high stringency to a
XX CC nucleic acid expressed in the human breast. The probes are useful for
XX CC predicting, diagnosing, grading, staging, monitoring and prognosing
XX CC diseases of the human breast, particularly those diseases with polygenic
XX CC aetiology. The diseases include: breast cancer, disorders of developmental
XX CC inflammatory diseases of the breast, fibrocystic changes, proliferative
XX CC breast disease and non-carcinoma tumours.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 312 AA;

Query Match 44.0%; Score 1325.5; DB 22; Length 312;
Best Local Similarity 93.7%; Pred. No. 2.5e-120;
Matches 254; Conservative 6; Mismatches 10; Indels 1; Gaps 1;

Qy 277 MLVGVNIVIKPASOTITCENCRLFTCIDSTFNWQHRIILVRAREGMWIPVSTDRPWEASP 336
Db 1 mlvvgnivikpdsqtqtcencrlltcidstfnwqhrillvraregvwllvsmndrpweasp 60

Qy 337 SIHILTEILKGVNRSKRFIPTLIIVIMGLIATVATAAGVALHSSVQSVNFVWYWK 396
Db 61 svhiltelkvgvnrskrfiftliavimgliatvataagvalhssvqsvnfvndwqkn 120

Qy 397 STRLWNSOSSIDQKLASQINDLRQTIVWMDRL-DLEHHFQLQCDWNTSDFCITPQIYNE 455
Db 121 strlwnsqssidqkqlangindlrqtivwmdrflmslehnrfqlqcdwntsdftcpqiye 180

Qy 456 SEHHWDMVRRHLQGRDNLTLDISKLEQIFEASKAHNLNVPGETAAGVADGLANLPV 515
Db 181 sehhdwmvrrhlqgrdnltldiskleqifeaskahlnlvpgteaagvadglaanlpv 240

Qy 516 TWIKTIRSTMIINLILVVCFLLLVCRCT 546
Db 241 twktigtsttiinlililvcfllllvcrct 271

RESULT 6
AAW95692
ID AAW95692 standard; Protein; 153 AA.
XX AC AAW95692;
XX DT 08-JUN-1999 (first entry)
XX DE Human endogenous retrovirus IDDMK1.2-22 env protein.
XX KW IDDM; insulin-dependent diabetes mellitus; endogenous retrovirus;
XX KW SAG; superantigen; provirus; autoimmune disease; type 1 diabetes;
XX KW diagnosis; env; envelope.
XX OS Homo sapiens.
XX PN EP893691-A1.

```

```

XX PD 27-JAN-1999.
XX XX
XX PF 23-JUL-1997; 97EP-0401773.
XX XX
XX PR 23-JUL-1997; 97EP-0401773.
XX XX
XX PA (MACH/) MACH B F.
XX XX
XX PI Conrad B, Mach B;
XX XX
XX DR WPI: 1999-097928/09.
XX DR N-PSDB; AAX07514.
XX XX
XX PT Diagnosing human autoimmune disease by detecting retrovirus with
XX PT superantigen activity - new retrovirus associated with type 1
XX PT diabetes, its proviral DNA, and related vectors, transformed cells,
XX PT proteins, antibodies and specific binding agents, used for treating
XX PT or preventing autoimmune disease
XX XX
XX PS Claim 31; Fig 7D; 92pp; English.
XX XX
XX CC The sequence is that of an insulin-dependent diabetes mellitus
XX CC associated human endogenous retrovirus (IDMK1.2-22) env protein.
XX CC The retrovirus has Superantigen (SAG) activity. It can be used
XX CC as part of a method is specifically used to diagnose type 1 diabetes
XX CC mellitus. Modified proteins expressed by the retroviral sequence
XX CC (without SAG activity but still able to induce an immune response)
XX CC are useful in vaccines to treat or prevent SAG-related autoimmune
XX CC disease; nucleic acid sequences encoding (modified) SAG can be used
XX CC similarly to treat such diseases. Retroviral-encoded SAG are important
XX CC in pathogenesis of autoimmune disease, probably by activating
XX CC autoreactive T cells. The method is very specific (it can differentiate
XX CC between expressed and non-expressed viral nucleic acids) and can be used
XX CC even where the pathogen is an ubiquitous endogenous retrovirus. Blood
XX CC or plasma samples can be tested without extensive preparation and
XX CC diagnosis can be made before clinical signs are apparent, allowing
XX CC early intervention before severe tissue damage has occurred.
XX XX
XX SQ Sequence 153 AA;

Query Match 27.9%; Score 840; DB 20; Length 153;
Best Local Similarity 100.0%; Pred. No. 1.4e-73;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVTPTVMDNPIEVYVNDVSVWPGPTDDRCAPKPEEGMMINISIGYHYPPICIGRAPGC 60
Db 1 mvtptvmdnpielyvndsvwvpgptddrcapkkeegmmnisigyhyppicigrpgc 60

Qy 61 LMPAVQNLWLVPTVSPNSRFTYHMGMSLRPRVNYLQDFSYORSKFRPKGTCPKKEI 120
Db 61 lmpavqnlwlvptvpsnsrftyhmvgmslrprvnylqdfsyqrslkfrpkgtcpkei 120

Qy 121 PKGSKNTEVLVWEECVANSVVLQNNFEGTIID 153
Db 121 pkgskntevlvweecvansvvlqnnfegtliid 153

RESULT 7
AAW97745
ID AAW97745 standard; Protein; 153 AA.
XX
XX AC AAW97745;
XX XX
XX DT 21-MAY-1999 (first entry)
XX XX
XX DE Human endogenous retrovirus IDDKK1.2-22 envelope protein.
XX KW HERV; IDDKK1.2-22; superantigen; SAG; antigen; IDDM;
XX KW insulin-dependent diabetes mellitus; autoimmune disease; diagnosis;
XX KW therapy; vaccine; envelope protein; env gene.
XX XX

```

```

OS Human endogenous retrovirus.
XX XX
XX FH Key Location/Qualifiers
XX FT Modified-site 17..19 "Asn is N-glycosylated"
XX FT Modified-site 42..45 "Asn is N-glycosylated"
XX FT
XX PN WO9905527-A2.
XX XX
XX PD 04-FEB-1999.
XX XX
XX PF 22-JUL-1998; 98WO-EP04926.
XX XX
XX PR 23-JUL-1997; 97EP-0401773.
XX PR 22-JUL-1997; 97EP-0112482.
XX XX
XX PA (MEDI-) MEDIGEN SA.
XX XX
XX PI Conrad B, Mach B;
XX XX
XX DR WPI: 1999-143118/12.
XX DR N-PSDB; AAX07188; AAX07189.
XX XX
XX PT New isolated human endogenous retrovirus - used to develop products
XX PT for the diagnosis, prevention and treatment of autoimmune disease,
XX PT particularly insulin dependent diabetes mellitus
XX XX
XX PS Claim 6; Fig 7D; 165pp; English.
XX XX
XX CC This is the envelope protein (Env) of a new human endogenous
XX CC retrovirus (HERV), designated IDDMK1.2-22, that has been identified
XX CC as the source of superantigen (SAG) activity in insulin-dependent
XX CC diabetes mellitus (IDDM) patients. The endogenous retrovirus is
XX CC ubiquitous in the human genome but is only expressed in diabetic
XX CC individuals. The HERV encodes SAG activity within the env gene.
XX CC A claimed process for the diagnosis, including the pre-symptomatic
XX CC diagnosis, of a human autoimmune disease associated with a HERV
XX CC having SAG activity comprises specifically detecting in a
XX CC biological sample either: (a) the mRNA of an expressed HERV having
XX CC SAG activity (especially IDDMK1.2-22 5'LTR, 3' long terminal
XX CC repeat, env or pol); (b) a protein or peptide expressed by the HERV
XX CC (see AAW97745-48); (c) antibodies specific to the proteins expressed
XX CC by the HERV; or (d) SAG activity specifically associated with the
XX CC HERV. Products of the invention can be used to identify substances
XX CC capable of blocking transcription or translation of SAG-encoding
XX CC nucleic acid sequences, useful in therapy and/or prevention of
XX CC autoimmune disease associated with the SAG. A nucleic acid encoding
XX CC human retroviral SAG can be used as a DNA vaccine. Expression of
XX CC the endogenous SAG in IDDM suggests a general model according to
XX CC which self SAG-driven and systemic activation of autoreactive T
XX CC cells leads to organ-specific autoimmune disease.
XX XX
XX SQ Sequence 153 AA;

```

```

Query Match 27.9%; Score 840; DB 20; Length 153;
Best Local Similarity 100.0%; Pred. No. 1.4e-73;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVTPTVMDNPIEVYVNDVSVWPGPTDDRCAPKPEEGMMINISIGYHYPPICIGRAPGC 60
Db 1 mvtptvmdnpielyvndsvwvpgptddrcapkkeegmmnisigyhyppicigrpgc 60

Qy 61 LMPAVQNLWLVPTVSPNSRFTYHMGMSLRPRVNYLQDFSYORSKFRPKGTCPKKEI 120
Db 61 lmpavqnlwlvptvpsnsrftyhmvgmslrprvnylqdfsyqrslkfrpkgtcpkei 120

Qy 121 PKGSKNTEVLVWEECVANSVVLQNNFEGTIID 153
Db 121 pkgskntevlvweecvansvvlqnnfegtliid 153

```

```

RESULT 8
AAW95694
ID AAW95694 standard; Protein: 181 AA.
XX
AC AAW95694;
XX
DT 08-JUN-1999 (first entry)
XX
DE Human endogenous retrovirus IDDMK1.2-22 env/fs (sag) protein.
XX
KW IDDM; insulin-dependent diabetes mellitus; endogenous retrovirus;
KW SAG; superantigen; provirus; autoimmune disease; type 1 diabetes;
KW diagnosis; env; envelope.
XX
OS Homo sapiens.
XX
PN EP893691-A1.
XX
PD 27-JAN-1999.
XX
PF 23-JUL-1997; 97EP-0401773.
XX
PR 23-JUL-1997; 97EP-0401773.
XX
PA (MACH/) MACH B F.
XX
PI Conrad B, Mach B;
XX
DR WPI; 1999-097928/09.
DR N-PSDB; AAX07516.
XX
PT Diagnosing human autoimmune disease by detecting retrovirus with
PT superantigen activity - new retrovirus associated with type 1
PT diabetes, its proviral DNA, and related vectors, transformed cells,
PT proteins, antibodies and specific binding agents, used for treating
PT or preventing autoimmune disease
XX
PS Claim 31; Fig 7G; 92pp; English.
XX
CC The sequence is that of an insulin-dependent diabetes mellitus
CC associated human endogenous retrovirus (IDDMK1.2-22) env/fs protein.
CC
CC The retrovirus has Superantigen (SAG) activity. It can be used
CC as part of a method is specifically used to diagnose type 1 diabetes
CC mellitus. Modified proteins expressed by the retroviral sequence
CC (without SAG activity but still able to induce an immune response)
CC are useful in vaccines to treat or prevent SAG-related autoimmune
CC disease; nucleic acid sequences encoding (modified) SAG can be used
CC similarly to treat such diseases. Retroviral-encoded SAG are important
CC in pathogenesis of autoimmune disease, probably by activating
CC autoreactive T cells. The method is very specific (it can differentiate
CC between expressed and non-expressed viral nucleic acids) and can be used
CC even where the pathogen is an ubiquitous endogenous retrovirus. Blood
CC or plasma samples can be tested without extensive preparation and
CC diagnosis can be made before clinical signs are apparent, allowing
CC early intervention before severe tissue damage has occurred.
XX
SQ Sequence 181 AA;

Query Match 27.9%; Score 840; DB 20; Length 181;
Best Local Similarity 100.0%; Pred. No. 1.8e-73;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVTPTWMDNPLEVYVNDVSVVPGPTDRCAPKEEGMMINISIGYHYPPICLGRAPGC 60
|||||
Db 1 mvtptwmdnplevyvndsvvpgptddrcpakeegmmnisigyhyppiclgapgc 60
|||||

QY 61 LMPAVQNMLVEVPTVSPNSRRTYHVMVSGMSLRPRVNYLQDFSYQBSLKRPKGKTCPKREI 120
|||||
Db 61 lmpavqnmlvevptvspnsrftyhmvsgmslrprvnylqdfsyqbslkrpkgkcpkei 120
|||||

QY 121 PKGSKNTEVLWEECVANSVVLQNEFGTIIID 153
|||||

```

```

Db 121 pkgskntevlvweecvansvvlqnefgtliid 153

RESULT 9
AAW97747
ID AAW97747 standard; Protein: 181 AA.
XX
AC AAW97747;
XX
DT 21-MAY-1999 (first entry)
XX
DE Human endogenous retrovirus IDDMK1.2-22 Env/F-S (SAG).
XX
KW HERV; IDDMK1.2-22; superantigen; SAG; antigen; IDDM;
KW insulin-dependent diabetes mellitus; autoimmune disease; diagnosis;
KW therapy; vaccine; envelope protein; env gene.
XX
OS Human endogenous retrovirus.
XX
FH Key Location/Qualifiers
FT Modified-site 17..19
FT /note= "Asn is N-glycosylated"
FT Modified-site 42..45
FT /note= "Asn is N-glycosylated"
XX
PN WO9905527-A2.
XX
PD 04-FEB-1999.
XX
PF 22-JUL-1998; 98WO-EP04926.
XX
PR 23-JUL-1997; 97EP-0401773.
PR 22-JUL-1997; 97EP-0112482.
XX
PA (MEDI-) MEDIGEN SA.
XX
PI Conrad B, Mach B;
XX
DR WPI; 1999-143118/12.
DR N-PSDB; AAX07191.
XX
PT New isolated human endogenous retrovirus - used to develop products
PT for the diagnosis, prevention and treatment of autoimmune disease,
PT particularly insulin dependent diabetes mellitus
XX
PS Claim 6; Fig 7G; 165pp; English.
XX
CC This is the envelope (Env)/FS polypeptide of a new human endogenous
CC retrovirus (HERV), designated IDDMK1.2-22, that has been identified
CC as the source of superantigen (SAG) activity in insulin-dependent
CC diabetes mellitus (IDDM) patients. The endogenous retrovirus is
CC ubiquitous in the human genome but is only expressed in diabetic
CC individuals. The HERV encodes SAG activity within the env gene.
CC A claimed process for the diagnosis, including the pre-symptomatic
CC diagnosis, of a human autoimmune disease associated with a HERV
CC having SAG activity comprises specifically detecting in a
CC biological sample either: (a) the mRNA of an expressed HERV having
CC SAG activity (especially IDDMK1.2-22 5'LTR, 3' long terminal
CC repeat, env or pol); (b) a protein or peptide expressed by the HERV
CC (see AAW97745-48); (c) antibodies specifically to the proteins expressed
CC by the HERV; or (d) SAG activity specifically associated with the
CC HERV. Products of the invention can be used to identify substances
CC capable of blocking transcription or translation of SAG-encoding
CC nucleic acid sequences, useful in therapy and/or prevention of
CC autoimmune disease associated with the SAG. A nucleic acid encoding
CC human retroviral SAG can be used as a DNA vaccine. Expression of
CC the endogenous SAG in IDDM suggests a general model according to
CC which self SAG-driven and systemic activation of autoreactive T
CC cells leads to organ-specific autoimmune disease.
XX
SQ Sequence 181 AA;

```


